Refining the picture story exercise : towards a better understanding of hope, fear, and the achievement motive

Ramsay, Jonathan E.

2014

Ramsay, J. E. (2014). Refining the picture story exercise : towards a better understanding of hope, fear, and the achievement motive. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/64818>

<https://doi.org/10.32657/10356/64818>

**REFINING THE PICTURE STORY EXERCISE:
TOWARDS A BETTER UNDERSTANDING OF
HOPE, FEAR, AND THE ACHIEVEMENT MOTIVE**

JONATHAN E. RAMSAY

**A THESIS SUBMITTED TO THE
NANYANG TECHNOLOGICAL UNIVERSITY
IN FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY**

2014

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor Prof. Joyce S. Pang, without whom the writing of this thesis would have been impossible. It's widely accepted that the supervisor-student relationship is one of the most important factors in predicting success and enjoyment in postgraduate study, which makes me doubly fortunate to have worked with one of the very best. Her enthusiasm inspired me to push my research in interesting new directions; while her patience and understanding made sure I never lost perspective or became discouraged. I could not have asked for a better mentor.

I would also like to thank Prof. Ringo M. Ho for his insight and guidance throughout my four years of study, and especially during the planning and implementation of my cluster analysis. With so many people relying on his considerable statistical expertise, I'm extremely grateful that he consistently made time to answer my questions.

In addition, I would also like to thank my thesis committee for their advice and comments, which have greatly improved my research. I would also like to thank Ms. Tan Ser Hong for all her tireless work coding my data, as well as all my participants, whose creativity made the analytic process that much more interesting.

Heartfelt thanks also go to my parents, whose belief has always empowered me to try new things, and to never settle for something that I don't truly believe in.

Finally, I would like to thank my wife Melissa for her continued love and support. Being married to an aspiring academic is not easy, and yet she has always supported my decision to pursue further study and has kept faith in my ability to succeed. Her compassion and loyalty are gifts to be treasured, and treasure them I do.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS i

TABLE OF CONTENTS ii

LIST OF TABLES ix

LIST OF FIGURES x

LIST OF APPENDICES xi

ABSTRACT xii

1. Achievement Motivation.....1

1.1. Implicit and explicit achievement motivation..... 2

1.2. Hope of success and fear of failure..... 6

1.3. The picture story exercise. 8

1.4. Research overview. 11

2. The Pictures..... 13

2.1. The characteristics of effective picture cues. 13

2.1.1. Cue strength. 14

2.1.2. Cue ambiguity. 15

2.1.3. Universality and relevance. 16

2.2. The characteristics of effective picture sets. 18

2.2.1. Size..... 18

2.2.2. Extensivity.....	19
2.2.3. Set ambiguity.....	20
2.3. Picture cue selection past and present	24
2.3.1. Picture cue selection in the TAT	24
2.3.2. Picture cue selection in the McClelland et al. (1953) PSE.....	25
2.3.3. Picture cue selection in the Heckhausen (1963b) PSE.....	27
2.3.4. Picture cue selection to the present day.....	27
2.4. Picture cue selection in contemporary research: Problems and pitfalls.	29
2.4.1. Lack of independent validation.....	29
2.4.2. Declining validity due to datedness and bias.	30
2.4.3. Suitability for the assessment of HS and FF.	33
2.5. Recent attempts to standardize and improve picture cue selection.....	34
2.6. The present research.....	37
3. The Content Coding System	39
3.1. Content coding in the TAT.	39
3.2. Content coding in the PSE.	40
3.3. Measuring <i>n</i> Ach: The McClelland et al. (1953) system.	42
3.3.1. Development.	42
3.3.2. Structure and specification.....	44
3.3.3. Validation.	47
3.3.4. Weaknesses.	49

3.4. Measuring nAch: The Winter (1994) system. 50

 3.4.1. Development. 50

 3.4.2. Structure and specification. 51

 3.4.3. Validation. 52

 3.4.4. Weaknesses. 54

3.5. nAch, HS, and FF in the McClelland/Atkinson tradition..... 54

 3.5.1. The early days: A tripartite division. 56

 3.5.2. Towards a separation of motives: TAQ and RAM. 57

3.6. Early projective measures of FF..... 59

3.7. Measuring HS and FF: The Heckhausen (1963b) system. 62

 3.7.1. Structure and specification. 64

 3.7.2. Validation. 66

 3.7.3. Weaknesses. 68

3.8. Measuring HS and FF: The Pang (2006) system. 69

 3.8.1. Structure and specification. 70

 3.8.2. Weaknesses. 71

3.9. The need for a consolidated coding system. 72

3.10. The present research..... 73

STUDY 1

4. Study 1 Overview..... 75

 4.1. Hypotheses..... 77

- 4.1.1. Test-retest reliability..... 77
 - 4.1.2. Divergent validity..... 78
 - 4.1.3. Predictive validity. 78
- 5. Study 1A: Developing the Picture Set..... 80**
 - 5.1. Method..... 80
 - 5.1.1. Participants..... 80
 - 5.1.2. Design and procedure..... 81
 - 5.2. Results..... 87
 - 5.2.1. Diversity..... 87
 - 5.2.2. Cue strength and cue ambiguity. 88
 - 5.3 Discussion..... 88
- 6. Study 1B: Validating the Picture Set..... 92**
 - 6.1. Method..... 92
 - 6.1.1. Participants..... 92
 - 6.1.2. Design and procedure..... 92
 - 6.1.3. Materials. 95
 - 6.1.4. Validation tasks. 96
 - 6.2. Results..... 100
 - 6.2.1. Test-retest reliability..... 100
 - 6.2.2. Divergent validity..... 103
 - 6.2.3. Predictive validity. 104

6.3. Discussion..... 107

STUDY 2

7. Study 2: Overview 109

8. Study 2A: Developing the Coding System 113

8.1. Method..... 113

8.1.1. Participants..... 113

8.1.2. Design and procedure..... 113

8.1.3. Materials. 114

8.1.4. Analysis..... 115

8.2 Results..... 124

8.3 Discussion..... 131

9. Study 2B: Formalizing the Coding System 134

9.1. Defining the coding system..... 134

9.1.1. nAch/HS..... 134

9.1.2. FF. 137

9.2. Writing the coding system..... 138

10. Study 2C: Validating and Refining the Coding System 143

10.1. Method. 145

10.2. Results..... 145

10.2.1. Test-retest reliability. 145

10.2.2. Divergent validity..... 146

10.2.3. Convergent validity..... 147

10.2.4. Predictive validity..... 147

10.3. Discussion..... 151

GENERAL DISCUSSION

11. Summary of Findings..... 154

11.1 Study 1..... 154

11.2 Study 2..... 155

12. Picture Set Contributions..... 157

12.1. An overdue reappraisal..... 157

12.2. Improved relevance and decreased bias..... 158

12.3. Appropriateness for the assessment of HS and FF..... 159

12.4. Independent picture set validation..... 160

12.5. The importance of set ambiguity..... 161

13. Coding System Contributions..... 162

13.1. A pathway to integration..... 162

13.2. Incorporation of the distinction between HS and FF..... 164

13.3. Simplification of administration..... 164

13.4. Avoidance of motive contamination..... 165

14. Wider Contributions..... 166

15. Limitations and Future Directions..... 169

16. Conclusions 178

REFERENCES 180

APPENDICES 204

LIST OF TABLES

Table 1. Demographic and achievement setting comparisons for the pretest cohorts, new picture set variants, and comparison picture sets 83

Table 2. Descriptive statistics and intercorrelations of motive scores in the ambiguous condition..... 102

Table 3. Descriptive statistics and intercorrelations of motive scores in the unambiguous condition..... 103

Table 4. Descriptive statistics for validation task variables in studies 1B and 2C... 104

Table 5. Psychometric properties of the five coding system versions..... 148

Table 6. Descriptive statistics and intercorrelations for motive scores in study 2C. 151

LIST OF FIGURES

Figure 1. Position of McClelland et al.'s (1953) categories in the adjustive behavioral sequence 45

Figure 2. Achievement cue strength of the pretested images 89

Figure 3. Dendrogram displaying the results of the average linkage cluster analysis 125

Figure 4. Dendrogram displaying the results of the complete linkage cluster analysis 126

Figure 5. Clustering solution derived from the average linkage analysis 130

LIST OF APPENDICES

Appendix A. Constituent images for the ambiguous picture set	204
Appendix B. Constituent images for the unambiguous picture set	205
Appendix C. Constituent images for the comparison picture set.....	206
Appendix D. Simplified manual for the assessment of approach- and avoidance-related achievement motivation.....	207

ABSTRACT

Decades of research have established an important distinction between the approach-related and avoidance-related manifestations of the need for achievement (*n*Ach), termed Hope of Success (HS) and Fear of Failure (FF) respectively. However, while researchers in the field of explicit motives have placed the approach-avoidance distinction at the heart of their theories of motivated behavior, implicit motive researchers have found it difficult to adequately assess and characterize HS and FF due to methodological limitations inherent in the most widely-used method of implicit motive assessment: the picture story exercise (PSE). The present research addresses these issues by redeveloping the two central components of the PSE instrument—the picture stimuli and the associated coding system—specifically for the purposes of HS and FF assessment.

Study 1 sought to develop a picture set that improved both the validity and reliability of HS/FF assessment while also investigating the impact of a previously unstudied methodological consideration termed *set ambiguity*. Study 2 sought to consolidate existing content coding systems for PSE assessment of *n*Ach, HS, and FF into a single composite system that simplified administration while retaining the best features of its constituents. Study 2 also clarified the relationships between *n*Ach, HS, and FF as measured by contemporary PSE measures, establishing the equivalence of *n*Ach and HS while reinforcing the distinction between *n*Ach/HS and FF.

In study 1A, four rounds of pretesting identified eight contemporary images that were effective in arousing both HS and FF. In study 1B, the validity and reliability of two versions of this picture set, one which depicted only achievement-related activity and one which was more ambiguous in its motivational content, were compared to those of a

comparison picture set comprising frequently used images. The ambiguous version of the new picture set was found to exhibit better test-retest reliability than the unambiguous version, while HS and FF scores derived from the ambiguous version were found to be more predictive of theoretically relevant behaviors than scores derived from either the unambiguous version or the comparison picture set.

In study 2A, data gathered during study 1 were coded for *n*Ach, HS, and FF using the four most prevalent content coding systems. The resulting subcategory and motive scores were cluster analyzed, yielding a two factor structure that supported the HS-FF distinction while confirming the operational equivalence of *n*Ach and HS. This clustering solution was subsequently used to guide to the construction of a simplified, consolidated content coding system in study 2B. In study 2C, data collected during study 1 were recoded using the new consolidated system. *n*Ach/HS and FF scores derived from this new system were found to exhibit test-retest reliability, convergent validity, and predictive validity.

Together, the new picture set and consolidated coding system address many of the most significant methodological issues in PSE assessment, and provide a powerful new tool for researchers wishing to investigate the approach- and avoidance-related manifestations of the achievement motive.

1. Achievement Motivation

For as long as psychology has existed as a scientific discipline, there has been a continued fascination with achievement motivation (Weiner, 2013). This interest is apparent in some of psychology's earliest works. William James (1890) speculated as to links between self-evaluation and achievement strivings, while early 20th century researchers such as Narziss Ach (e.g., Ach, 1910) and Albert Michotte (e.g., Michotte & Prüm, 1910) pioneered the empirical study of achievement-related constructs such as effort and performance in their work on volition. As the embryonic field developed into different disciplines, this prominence continued, with achievement motivation taking center stage in early theories of motivation and personality psychology. Murray (1938) identified the need for achievement (grouped with several other needs under the broad, achievement-relevant domain of ambition) as one of the fundamental psychogenic needs in his theory of personology, and his work subsequently inspired the investigations of McClelland and Atkinson, two researchers whose work would solidify the central place of achievement motivation in scientific psychology (Elliot & Dweck, 2005).

The influence of these latter three researchers—Murray, McClelland, and Atkinson—cannot be overstated. Not only did they spearhead the systematic study of achievement motivation, they also developed and refined the most well-established methodology for measuring the achievement motive and related constructs, the Thematic Apperception Test (TAT) and its successor the Picture Story Exercise (PSE), which require respondents to write short stories in response to motivationally arousing images. To the uninitiated, this might sound like an odd way of measuring motivation. Contemporary personality research is dominated by the use of self-report questionnaires,

occasionally supplemented with alternate methods such as other-ratings and behavioral observation, and projective measures such as the TAT/PSE are often viewed skeptically or dismissed as unscientific relics from the days of leather couches and swinging pocket watches. Despite these negative appraisals, the PSE continues to be widely used in achievement motivation, and its enduring popularity can largely be attributed to the soundness of the rationale behind its initial development. Murray, and later McClelland and Atkinson, believed that many of the most significant motives, including the need for achievement, operate largely at an unconscious level, influencing behavior in ways that are imperceptible to the individual in question. As such, these motives are not accessible to conscious introspection and cannot be assessed using self-report methods. This belief was the fundamental guiding principle for their research and the work it went on to inspire (Schultheiss & Brunstein, 2010), and it necessitated the use of projective methods, in which the presence or strength of latent psychological constructs (such as motivation) are inferred from the content of speech or action. While this approach is not without its detractors (Entwistle, 1972), projective assessment continues to play a significant role in achievement motivation research to this day.

1.1. Implicit and explicit achievement motivation.

The need for achievement, along with the need for power and the need for affiliation, comprise the motivational trinity known as the *big three*, which represent the most common and significant human strivings or goals (Hofer & Chasiotis, 2004; Winter, 1994). While each of the big three motives has a rich history of empirical investigation, the need for achievement is undoubtedly the most widely studied, likely due to its central

importance to such significant domains as education and employment. The achievement motive, as well as closely-related constructs such as the need for competence (Deci & Ryan, 1985; Elliot, McGregor, & Thrash, 2002) and achievement goals (Dweck & Elliott, 1983; Elliot & McGregor, 2001), have been shown to positively influence such diverse outcomes as course interest and exam grades (Harackiewicz et al., 2000), goal-setting (Heckhausen, 1963b), entrepreneurial success (McClelland, 1987), and job performance (Baard, Deci, & Ryan, 2004). Given the above, it is perhaps unsurprising that psychologists in both theoretical and applied settings have expended considerable effort identifying the causes and effects of achievement motivation.

Pang (2010a, p. 30) defines the achievement motive as a “non-conscious and recurrent preference for affectively rewarding experiences related to improving one’s performance”. In highlighting the non-conscious nature of the achievement motive, this definition aligns closely with the conceptualizations of Murray, McClelland, and Atkinson, who adopted projective assessment techniques in order to render the unconscious amenable to investigation. Those working in this theoretical tradition speak of a fundamental distinction between two classes of motives: *implicit motives*, which operate outside of conscious awareness, and *explicit motives*, which can be accurately self-assessed, verbalized, and reflected-upon. In a seminal paper, McClelland, Koestner, and Weinberger (1989) summarized the key differences between implicit and explicit (or self-attributed) motivation. Implicit motive systems develop earlier in life and are automatized in the sense that they are activated unconsciously by affective incentives that are intrinsic to the situation or task. Explicit motives arise later in conjunction with

language development and are based around the construct of the social self. These motives are heavily linked to social expectation and the normative values of important social groups (e.g., family, peers, society at large), and are triggered by affective incentives linked to social evaluation, such as assessment or appraisal. Thus while implicit and explicit motive systems are both based in affect in the sense they both direct behavior in order to obtain anticipated reward (i.e., positive affect) and avoid anticipated punishment (i.e., negative affect), they differ fundamentally in terms of the classes of incentives that these anticipated affective experiences relate to. Consequently, explicit motives predict immediate, specific responses to socially incentivized situations (e.g., self-reports of aspiration and ability; Heckhausen & Halisch, 1986), whereas implicit motives predict spontaneous behaviors over time (e.g., entrepreneurship; McClelland, 1965).

This distinction yields two very different kinds of achievement motivation. The implicit need for achievement (*nAch*) represents a desire to excel in a task for its own sake, and to continually better one's performance in order to reap the affective rewards associated with proficiency and mastery. Implicit achievement motivation can be contrasted with explicit or self-attributed achievement motivation (*sanAch*), in which motivation to achieve stems instead from the social desirability of success and the social stigma attached to failure. *sanAch* motivated individuals value achievement highly because the societies in which they were raised place a positive incentive value on achievement. As a consequence, they strive to achieve (or to provide responses consistent with an achievement orientation) because they believe that social others will view them

more favorably, which in turn enhances self-esteem and positive affect. Patten and White (1977) observed that individuals scoring highly for *sanAch* perform significantly better than low scorers only in the presence of external incentives (in this case, participants were told that their scores on the task were indicative of leadership potential and general intelligence), whereas implicit achievement motivation predicts performance on tasks such as solving anagrams (Raynor & Entin, 1982) even in the absence of such incentives, with individuals striving to achieve for its own sake rather than to satisfy extrinsic criteria. A weight of empirical evidence supports this fundamental distinction. Spangler (1992) conducted a meta-analysis of PSE and self-report achievement motive research, investigating the utility of these two kinds of measures in predicting different classes of achievement-related behavior. In consolidating the findings of 105 studies, he found that the presence of task-related incentives strongly predicted behavior in PSE studies, whereas the presence of social incentives (e.g., rewards, expectations, demands, or norms associated with task performance) strongly predicted behavior in studies where the achievement motive was assessed by way of self-report.

The distinction between implicit and explicit achievement motivation has important consequences in the workplace. Given that implicit motives predict long-term behavioral trends (such as sustained productivity over time), rather than immediate responses to specific situations (such as productivity under direct supervision), accurate assessment of the implicit need for achievement, and the identification of those individuals who possess it to a significant degree, should be a top priority for all organizations. Since individuals who are implicitly motivated to achieve will strive to perform well even in the absence of extrinsic incentives, it is likely that such individuals

will have a greater positive impact on an organization's overall performance, and a wealth of evidence from the McClelland research tradition demonstrates the value of measures of implicit achievement motivation in predicting many positive work-related outcomes (McClelland, 1987).

1.2. Hope of success and fear of failure.

Another important distinction in achievement motivation is between the approach-related and avoidance-related manifestations of this motive, referred to as Hope of Success (HS) and Fear of Failure (FF) respectively. HS-motivated individuals are driven to engage in motive-relevant behavior by the affective incentives associated with successful task completion, whereas FF-motivated individuals are driven by the affective disincentives associated with failure. This distinction between approach and avoidance resonates with Higgins' (1998) theory of regulatory focus, which states that individuals vary in terms of whether they prioritize the pursuit of reward (promotion focus) or the avoidance of punishment (prevention focus). Promotion-focused individuals exhibit sensitivity to the presence or absence of positive outcomes, use approach-related strategic means, seek to maximize gains, and experience changes in cheerfulness/dejection-related emotions in response to motive-relevant outcomes. On the other hand, those with a prevention focus are more sensitive to the presence or absence of negative outcomes, use avoidant behavioral strategies, seek to minimize losses, and experience changes in quiescence/agitation-related emotions in response to success or failure. This contrast between approach focus and avoidance focus has proved to be highly influential, and many motivational theories incorporate this distinction (e.g., Carver, 2006; Elliot &

Church, 1997; Roth & Cohen, 1986), with some theorists claiming that the approach-avoidance distinction is “fundamental and basic, and should be construed as the foundation on which other motivational distinctions rest” (Elliot & Covington, 2001, p. 74).

Traditionally, HS has been conceived of as the ideal form of achievement motivation (indeed, it is frequently seen as being analogous to *nAch*), engendering heightened engagement and persistence (Feather, 1962), superior performance (Lowell, 1952), and better psychological outcomes such as task enjoyment (Puca & Schmalt, 1999), whereas FF, which is thought to be rooted in anxiety, reduces self-efficacy and inhibits the expression of achievement-related behaviors, worsening performance and enhancing negative affect (Covington & Omelich, 1985). However, the work of Heinz Heckhausen, another of the field’s most influential figures, pointed to a more complex alternative scenario. During the 1960s, Heckhausen modified the PSE in order to simultaneously examine both HS and FF, obtaining results that suggested that FF, like HS, may also be activating with respect to achievement behavior (see section 3.7). Despite these advances, the implicit motive research community remained fixated on the simpler conceptualization for much of the 20th century, and it is only in the last two decades that Heckhausen’s alternative characterization of FF, and PSE assessment of FF in general, have experienced a limited resurgence.

The distinction between HS and FF also has important practical implications. While work in the Heckhausen tradition has demonstrated that FF is activating with respect to achievement motivation, there can be little doubt that avoidant achievement motivation is associated with a host of negative outcomes in both the workplace and the

classroom. Findings from previous research (Heckhausen, 1980) suggest that the performance of FF-motivated individuals suffers when they're placed under time pressure, meaning that highly time-pressured jobs such as paramedic or newspaper journalist may be less suitable for someone with this achievement orientation. Similarly, avoidance-related achievement goals, which are a logical counterpart to the avoidant achievement drive that is fear of failure, have been found to predict a variety of suboptimal academic outcomes. Elliot et al. (1999) found that adoption of avoidant achievement goals was associated with superficial processing and tendencies towards disorganization, such as an inefficient use of study time, which subsequently predicted decreases in academic performance. Similarly, Bouffard et al. (1995) found that avoidance-oriented college students displayed a pattern of reduced effort and persistence, possibly as self-handicapping strategies (Berglas & Jones, 1978) that will mitigate the impact of potential failure on self-esteem. More generally, Roney et al. (1995) found that avoidant or prevention-focused individuals are less persistent and perform worse than promotion focused individuals when attempting to solve impossible anagrams. These findings suggest that the academic performance of FF-motivated students may be continually undermined by various behavioral tendencies (e.g., procrastination, giving up) that limit their potential. Given this situation, the importance of identifying and assisting failure-oriented students and employees is obvious.

1.3. The picture story exercise.

Implicit motive researchers continue to make widespread use of the PSE in their assessment of the achievement motive. While alternative measures have been developed,

such as the multimotive grid (MMG; Sokolowski, Schmalt, Langens, & Puca, 2000) and the operant motive test (OMT; Kuhl & Scheffer, 1999), none can claim the same wealth of validation data as the PSE, and recent research reporting low intercorrelations between these resulting motive scores has cast doubt on whether these measures assess the same underlying constructs (Schüler, Wegner, & Brandstätter, 2013). Consequently, the PSE remains the default tool for the assessment of implicit achievement motivation, despite its reputation as a labor-intensive methodology that is difficult to master (Weinberger, Cotler, & Fishman, 2010) and despite criticisms of its psychometric adequacy (e.g., Entwistle, 1972; for a rebuttal of these criticisms see Pang, 2010b).

The picture story exercise is an instrument with two distinct components: a picture set and a content coding system. When completing the PSE, a respondent is shown a sequence of picture stimuli, individually referred to as cues, which depict one or more individuals engaged in motive-relevant behaviors. Images in the picture set are presented one at a time, usually for only a brief period of around ten seconds (Pang, 2010b), after which participants are given a limited amount of time (typically between four and five minutes; Smith et al. 1992) to write an imaginative story inspired by the picture they just saw. Only limited instructions are provided, with respondents being encouraged to write stories with a defined narrative structure (i.e., “a beginning, a middle, and an end”; Pang, 2010b, p. 135), and that emphasize the internal states of the actors involved (“What are the people thinking about and feeling? What do they want?”; Schultheiss & Pang, 2007). Once the stories have been written, trained PSE researchers use content coding systems to identify the presence of motive-relevant imagery. These content-coding systems are

essentially rule books: carefully worded manuals that outline what qualifies as motive imagery and what doesn't. They are typically organized in a categorical fashion, presenting the coder with a series of discrete classes of imagery relating to different manifestations of the motive. During analysis, coders read the stories and note down each instance of motive imagery, summing to give individual picture scores that are then averaged across the entire picture set in accordance with the rules of that particular coding system. These finalized composites are the ultimate product of PSE assessment: numerical scores that are taken to be indicative of underlying motivational traits or arousal states.

Despite the PSE's enduring popularity, there are several limitations that hinder its use. These methodological issues relate to both the picture set and the coding system components, hampering efforts to integrate findings from different traditions, compromising the validity of assessment, and perhaps most damagingly of all, preventing researchers in the field of implicit motivation from fully incorporating the HS/FF distinction into their work. The primary aim of the present research is to provide a comprehensive solution to these problems, placing equal emphasis on both the picture set and the coding system. Despite tacit acknowledgment that the picture set and coding system are of equal importance (Pang, 2010b), there has been a far greater emphasis on the content coding systems in research (Schultheiss & Brunstein, 2001), a historical trend that undervalues the critical role of the picture set. I hope that the methods adopted in this thesis will go some way to redressing this imbalance.

1.4. Research overview.

The first part of this introduction will discuss the PSE's frequently under-appreciated component: the picture set. The discussion will commence with an overview of current guidelines regarding picture cue selection, first describing the properties of effective picture cues (e.g., cue strength, cue ambiguity) and of effective picture sets (e.g., size, extensity), before proposing a previously undocumented consideration: set ambiguity. This discussion will preface a chronological account of picture cue selection methods over the years, beginning with picture cue selection in the TAT before progressing to the novel criteria employed in the McClelland/Atkinson and Heckhausen traditions. Subsequently, it will be argued that commonly used picture cues may suffer from declining validity, and that insufficient effort has been made to develop new picture sets since PSE's genesis in the mid-20th century. Furthermore, it will be argued that the majority of commonly used pictures may be inappropriate for the assessment of HS and FF, and that this is a significant contributing factor to the marginalization of this distinction in implicit motive research. After highlighting the problems associated with contemporary picture sets, the introduction will outline a rationale for the development and validation of a novel picture set intended to address these concerns. This is the first major aim of the present research.

The second part of the introduction will argue for the necessity of improving and standardizing the content coding systems used in PSE assessment of the achievement motive. First, a historical perspective on the development of the content coding systems that both define and measure the achievement motive will be provided, commencing with

the PSE's predecessor, the TAT (Morgan & Murray, 1935), before going on to consider the McClelland, Atkinson, Clark, and Lowell (1953), Winter (1994), Heckhausen (1963b), and Pang (2006) content coding systems. The strengths and weaknesses of each system will then be identified, and the theoretical disparities between these approaches, which have contributed significantly to the fragmentation of the field and the tendency to ignore the HS/FF distinction, will be highlighted. It will then be argued that a consolidated system, one which incorporates the distinction between approach and avoidance while drawing on the other strengths of existing measures, represents a promising solution to the problems currently facing researchers in the field of implicit achievement motivation. Development and validation of such a system is the second major aim of the present research.

2. The Pictures

Despite extensive efforts to develop reliable and valid content coding systems, the stimuli—the picture cues themselves—remain the most fundamental component of the PSE. It is therefore strange and unfortunate that the methodological considerations of picture cue selection have received comparatively little attention in the literature.

Schultheiss and Brunstein (2001) observed that PSE researchers frequently “rely on their intuition or on personal advice by other scholars when deciding what picture cues to use in assessing a particular motive or combination of motives” (p. 72), when such decisions should rightfully be guided by empirically derived procedures and guidelines. While there has been a recent reawakening of interest in the study of picture sets (e.g., Pang, 2010b; Pang & Schultheiss, 2005; Schultheiss & Pang, 2007), our understanding of picture sets continues to lag behind that of content coding systems, and prominent voices in the field have called for this situation to be rectified (e.g., Schultheiss & Brunstein, 2001).

2.1. The characteristics of effective picture cues.

The importance of the picture cues to the overall function of the PSE cannot be overstated. As “major determinants of thematic apperceptive story content” (Smith, Feld, & Franz, 1992, p. 517), picture cues light the fuse of implicit motive arousal, subconsciously prompting respondents to project motive-relevant imagery onto the characters and situations they depict. The central assumption of the PSE, and indeed of all projective techniques, is that the intensity of this projection reflects the strength of the underlying dispositional motive: the centrality of this unconscious desire to the individual's psychological make-up. Thus in order for the PSE to function as a measure of

implicit motivation, the constituent images must be effective in stimulating varying degrees of projection among motivationally diverse individuals. If the images are too effective, a majority of the resulting stories will feature motive-relevant imagery. If they are largely ineffective, the amount of codable imagery will be negligible. In both cases the resulting homogeneity of generated content renders the PSE useless, as the variability of the outcome measure (motive image frequency) is compressed to such an extent that it no longer captures individual differences in motive strength. This need for appropriate balance was acknowledged by early researchers in the McClelland tradition (e.g., Birney, 1958; Haber & Alpert, 1958; Murstein, 1965a; 1965b), who strove to identify the various properties that effective picture cues should possess.

2.1.1. Cue strength.

The foremost and perhaps the most intuitively obvious of these properties is *cue strength*: the ability of a given picture cue to elicit motive-relevant imagery. Cue strength is typically gauged by measuring *motive pull*, which is defined as the frequency with which the image elicits codable motive imagery, and given the preceding observation that picture cues must prompt varying degrees of motive projection amongst different individuals, it follows that a picture cue must possess sufficient pull for the target motive in order to provide a valid measure of the construct. While the results of McClelland, Clark, Roby, and Atkinson's (1949) pioneering study, which was the first application of the PSE to the achievement motive, initially suggested that strong and weak images discriminated equally well between high and low *nAch* scorers (McClelland et al., 1953, p198), contemporary researchers generally advise choosing images that display medium-high pull for the motive under investigation (e.g., Pang, 2010b; Schultheiss & Brunstein,

2001), with images being defined as high pull when at least 50% of participants respond with at least one instance of codable imagery. Utilizing images with either excessive or insufficient pull is problematic, not only because it reduces inter-story variability and compromises validity, but also because it gives rise to highly skewed motive score distributions that are not amenable to traditional statistical analysis (Schultheiss & Pang, 2007).

2.1.2. *Cue ambiguity.*

Cue strength is closely related to another important property: *cue ambiguity*.

Whereas cue strength represents the ability of an image to elicit imagery corresponding to a single motive, cue ambiguity is defined as the tendency to elicit imagery pertaining to multiple motives. While an unambiguous image can pull strongly for either *nAch* or *nAff* (but not both), an ambiguous image may pull for these two motives simultaneously.

Strong pull for multiple motives is relatively rare, as an increase in ambiguity is typically accompanied by a concomitant decrease in pull for the various motives (Pang, 2010b). As with the issue of cue strength, accepted protocol with regard to selection of ambiguous versus unambiguous picture cues has changed somewhat over the years. Early researchers suggested eschewing ambiguous images on the grounds that they reduce the PSE's effectiveness. Haber and Alpert (1958) observed a significant negative correlation between cue ambiguity (operationalized as subjective assessment of ambiguity by participants) and cue strength, as well as decreased test-retest reliability when using ambiguous versus unambiguous cues. However, more recent work suggests that a degree of cue ambiguity is not only desirable but also necessary. Pang (2010b) advises researchers to select picture cues with sufficient cue ambiguity for two reasons. Firstly,

because cue ambiguity is typically inversely related to cue strength, choosing ambiguous images ensures that resulting motive scores are not artificially compressed (and hence invalidated), which may be the case when high pull images give rise to stories saturated with motivational content. Cue ambiguity is therefore advantageous as it increases variance in resulting motive scores. This possibility has been suggested by Murstein (1965a), who observed that “the high cards had such an overwhelming pull for *nAch*, that subjects responded with achievement regardless of their own *nAch*” (p. 286). Secondly, Pang (2010b) cites an intriguing finding by Clark and Sensibar (1955), which suggests that the use of unambiguous images may have unintended adverse consequences for motive assessment. Clark and Sensibar (1955) found significantly more sexual content in stories written by male undergraduates in response to mundane landscape and architecture images than in comparable stories written in response to more risqué images depicting women in varying states of undress. Pang (2010b) suggests that overtly sexual nature of the pornographic images may have aroused suspicion among participants, leading to intentional motive suppression and the deliberate omission of carnal imagery. This possibility would suggest that highly unambiguous images risk exposing the experimenter's intentions, and thus undermining the necessarily covert administration of the PSE.

2.1.3. Universality and relevance.

Cue strength is considered to be a function of several image properties. In addition to cue ambiguity, which concerns how obviously a given image depicts target motive-relevant activity (compared to that of other motives), cue strength is known to be influenced by both the *universality* and *relevance* of the image: two properties that dictate

the likelihood of motive projection by the respondent. An image can be considered truly universal if it holds identical motivational significance for every member of the target population, and while achieving true universality is impossible given the vastly different activities and situations associated with motive satisfaction by different individuals, images should be chosen in order to maximize universality since this assures a minimum level of motivational pull across participants. Establishment of such a baseline allows for more accurate detection of differences in motive tendencies (Atkinson, 1958). Related to universality is relevance, which is the extent to which an image captures the concerns and experiences of the average population member. If the depicted scene is relevant and hence meaningful to the respondent, there is a greater likelihood that exposure to the stimulus will spark a motivationally-relevant train of thought. Although conceptually similar, universality and relevance are not the same. The frequently used image *trapeze artists* (McClelland, 1975) provides an illustrative example. While the notion of circus acrobatics may be widely and similarly understood (high universality), it seems safe to assume that such an image does not represent the current concerns of the average population member (low relevance). Since increased universality and relevance boost cue strength, researchers in the field recommend choosing PSE images that are both universal and relevant. Smith et al. (1992) advise choosing pictures “whose content constitutes a reasonably representative sample of motive-relevant situations that have similar significance to all members of the population to be tested” (p. 519), while Pang (2010b) advocates the use of images depicting scenarios in which the motive of interest is typically aroused, such as work, school, and competitive sports in the case of *nAch*.

2.2. The characteristics of effective picture sets.

If the desirable properties of PSE picture cues have received only limited attention, then the properties of overall picture sets have received less still. This is an unfortunate situation, since the relationships between cues and the overall composition of the picture set may be just as important as the properties of the individual cues. Although neglected for many years, systematic efforts have recently been made to identify the properties of picture sets that contribute to the reliability and validity of the PSE. Some properties, such as picture set size and extensivity, have been found to make critical contributions to PSE validity, while others, such as presentation order, have been found to be of minimal importance (Pang & Schultheiss, 2005).

2.2.1. Size.

According to classical test theory, increasing the number of items in a self-report scale improves the measure's psychometric properties, since combined scores derived from large numbers of items typically provide a better approximation of the underlying latent variable. In the case of the PSE however, increasing the number of items (i.e., the number of pictures in the set) does not offer comparable benefits. Given the effortful and involved nature of writing imaginative stories, asking participants to write too many stories is likely to cause fatigue, and runs the risk of causing annoyance or disengagement. In their 1958 investigation into the effects of picture serial position, Reitman and Atkinson found motive scores derived from pictures in the second half of a set were considerably less predictive of theoretically-relevant behaviors than scores derived from first half pictures. This observation of declining validity as participants progress through the picture set prompted them to question how large a picture set can become before scores derived

from the latter pictures begin to become invalid. Even within a set of eight pictures, Reitman and Atkinson noticed a decline in validity in the second half, leading them to speculate that some participants were adapting to the procedure in some way (p. 682), although they did not speculate as to the nature or cause of this adaptation process. On the other hand, Schultheiss and Pang (2007) caution against the usage of picture sets with too few images, since reducing the number of pictures logically reduces the variability of the resulting scores, thus compromising validity. Furthermore, they provide evidence that the use of four or fewer picture cues gives rise to motive score distributions with high positive skew, due to the high number of individuals showing zero motive imagery. Such distributions are not amenable to linear regression even after square root or logarithmic transformation, reducing the scope of potential analysis. In addition to these limitations, it is also contended that reducing the number of pictures necessarily causes a decrease in the picture set's extensity, which can also compromise the validity of the resulting scores (see section 2.2.2 for more details).

2.2.2. Extensity.

The variability of the domains and situations in which different individuals seek satisfaction of the same underlying motive mean that picture sets can also vary in terms of *extensity*, defined as the degree to which a picture set features a diverse range of motivationally relevant scenarios. Since individuals tend only to project onto characters and situations that are personally relevant (see section 2.1.3.), a varied picture set will be of greater relevance to a wider range of people, and will thus be more likely to elicit codable motive imagery. A picture set comprised entirely of academic or scholastic images, such as students poring over text books or attending lectures, would be said to

exhibit low extensity, while a picture set that balances academic settings with sporting or employment contexts would be said to exhibit relatively high extensity. Pang and Schultheiss (2007) recommend using picture sets with high extensity particularly when studying the relationship between implicit motives and global variables (such as emotional well-being over time or life outcomes), while suggesting that picture sets with lower extensity can be utilized when the content of the picture cues closely matches the outcome variable under investigation (e.g., using picture cues depicting debate or argument when assessing the impact of power motivation on persuasive communication). Although the lack of systematic studies of extensity has been noted (Pang, 2010b), researchers generally recommend the use of picture sets with high extensity for both single motive and multi-motive assessment (see also Pang 2010b, Smith et al., 1992).

2.2.3. *Set ambiguity.*

Despite these advances, there is still much to learn with respect to optimizing picture set composition. For example, established protocol has been somewhat hazy on the issue of whether picture sets should include a variety of images that pull selectively (i.e., they exhibit moderate-high cue strength for a target motive and low cue ambiguity) for different motives, or whether picture cue batteries should be more uniform in their make-up. In the absence of existing terminology, this variable will hereafter be referred to as *set ambiguity*. A picture set comprising images that pull for different target motives exhibits set ambiguity, while a picture set containing only *nAch*, *nAff*, or *nPow* images does not. Set ambiguity should not be confused with cue ambiguity, which is discussed in section 2.1.2. Cue ambiguity refers to the ability of individual pictures to pull for several motives simultaneously, while set ambiguity refers to the inclusion of images of varying

cue strengths and motivational targets in a given picture set. Both are related to differences in pull for different motives, yet they differ in terms of their levels of analysis: individual images in the case of cue ambiguity, and combined picture sets in the case of set ambiguity.

Generally, researchers advise that picture sets should be ambiguous when studying different motives (Smith et al., 1992; Veroff et al., 1960). This is sensible advice, since the administration of images with sufficient pull for each of the target motives is necessary for valid assessment. The prevailing wisdom differs for single motive study. Schultheiss and Pang (2007) advise researchers assessing a single motive to “select the four to six pictures which have the strongest pull for that motive” (p. 331), while both Smith et al. (1992) and Pang (2010b) encourage researchers to include only moderate-high pull images for the target motive. There is however no empirical evidence to substantiate claims that unambiguous picture sets result in more valid assessment, and there is reason to suspect that including only images with strong target motive pull may be counter-productive.

Firstly, unambiguous picture sets may suffer from decreased validity for the same reasons that unambiguous picture cues do. More ambiguous picture sets including items that pull preferentially for alternate motives may give rise to more valid scores because of increased response variability, which in turn makes it easier to distinguish high and low target motive scorers. For example, a picture set that exclusively comprises images that pull primarily for *n*Ach may produce similar scores for both highly and moderately achievement motivated individuals, whereas only the highly motivated individual is likely to project achievement motivation onto a picture set containing a number of images that pull preferentially for alternate motives. As such, it is possible that unambiguous picture

sets would give rise to a contracted range of scores, borne of uniform responding to achievement-related content, that do not accurately reflect the true motive variability of the sample. Such scores would display reduced predictive validity, since the ability of an independent variable to predict an outcome is necessarily reduced as its variability is artificially compressed.

Secondly, Atkinson and Birch's (1970) dynamics of action (DOA) theory suggests that writing motivationally-relevant stories possesses consummatory value, in that expressing the underlying need is motivationally satisfying in and of itself, reducing the tendency for further expression. This theory was used to explain both the "sawtooth" pattern of rising and falling motivational strength over time and also the low internal consistency of PSE implicit motives measures (Atkinson, 1992). According to DOA theory (see Atkinson, 1992, pp. 26-28), the tendency to write motive-relevant imagery (T) is a function of the instigating force (F; i.e., the underlying motive) and the consummatory force (C; the motivational satisfaction derived from expression). When responding to uniform, unambiguous picture sets, C remains consistently high, decreasing the tendency to write motive-relevant imagery and reducing the frequency of the aforementioned motivational wave. Over the course of a picture set, typically including between five and eight images, this method may under-represent the intensity of the motive disposition, since measurement points will more likely fall in these amotivational troughs where consummatory force is high. More ambiguous picture sets may offer improvement, since the inclusion of images that arouse alternate motives enforces breaks during which target motive-related C can subside, increasing the tendency to write target motive-related imagery in response to subsequent pictures. As such, ambiguous picture sets may be less

likely to underestimate the true values of implicit motives.

Thirdly, responses to funneled debriefing questions in previous PSE studies I have conducted suggested a degree of hypothesis awareness among participants responding to unambiguous picture sets. After completing an unambiguous achievement PSE some participants indicated that they believed that they were being asked to write stories about success and failure in achievement tasks. This revelation is problematic, because if participants are altering their responses due to demand characteristics, then the PSE becomes more of a measure of explicit motivation, since participants are in effect endorsing self-referenced motivational statements.

There is also reason to suspect that unambiguous picture sets may also give rise to less reliable scores. This possibility may appear to be somewhat counter-intuitive. Given the previous assertion that unambiguous picture sets may lead to reduced response variability, it might actually be expected that unambiguous picture sets give rise to scores that are more stable over time despite being less valid. However, researchers (e.g., Schultheiss & Pang, 2007) have noted that PSE respondents frequently feel compelled to be original and change their stories when responding to pictures for a second time, thus reducing test-retest reliability. It is therefore contended that this tendency may be exaggerated in individuals responding to successive administrations of an unambiguous picture set, since the similarity in content of the stories written at time one should heighten the need to introduce variety at time two, thus further reducing test-retest reliability.

The points raised previously speak of a need to directly assess the role of set ambiguity in determining the validity and reliability of the picture story exercise. This

need, and the methods proposed to fulfill it, will be discussed in more detail in section 2.6.

2.3. Picture cue selection past and present.

Despite the formulation of picture selection guidelines by Pang (2010b) and Smith et al. (1992), both of whom explicitly encourage the selection and pre-testing of novel picture cues, PSE researchers have generally been reluctant to stray from the beaten track. Since striking an appropriate balance between cue strength and cue ambiguity is critical to ensuring the validity of the PSE's resulting scores, this reluctance is to some extent understandable. Utilizing previously validated picture cues is efficient, since the researcher need not concern herself with the labors of pre-testing for suitability, and methodologically appropriate, since the ability of these images (at least in partnership with their associated coding systems) to produce valid scores with adequate variance has already been established. Even a casual survey of the recent literature clearly illustrates this unwillingness to reinvent the wheel, and many contemporary researchers have noted the reliance on older, previously validated images (Langen-Fox & Grant, 2006; Schultheiss & Brunstein, 2001).

2.3.1. Picture cue selection in the TAT.

In the PSE's progenitor, the TAT (Morgan & Murray, 1935), images were designed to be evocative of specific psychodynamic issues or processes. Murray, who had developed a keen interest in the diagnostic power of projection (which, unlike Freud, he viewed as a generalized tendency to project one's internal states onto external agents or objects, not a highly specific ego defense mechanism), initially showed photographs of male faces to children before and after playing a frightening game, finding that depicted

individuals were judged to be significantly more malicious after playing the game (Murray, 1933; cited in Anderson, 1999). Such preliminary experiments convinced Murray that pictorial cues effectively stimulated projection, affording him insight into his participants' subconscious states, a conviction that was to be the guiding principle in his co-development of the TAT. As he sought to formalize his projection-induction procedure into a clinical tool, he realized the value of developing a standardized picture set that could be given to a variety of participants, and consequently went about selecting the most effective cues from a collection of images he had amassed over the course of his work. These pictures were selected on the basis of four factors, although it is likely that Murray's own preferences and biases also influenced the decision-making process (Anderson, 1999), a point that, in retrospect, underscores the need for more objective criteria. These four factors bear a striking resemblance to contemporary cue selection criteria. Comprehensiveness refers to the depiction of a wide range of scenarios and relationships within the picture set (conceptually similar to the notion of extensity), while ambiguity refers to the precisely the same property: the avoidance of images that "told too much" (Anderson, 1999, p. 31) and hence generated uniform narratives. Identification concerned whether the respondent would identify with any of the depicted characters and so captures the essence of both universality and relevance. The final criterion, stimulating power, is theoretically identical to the notion of cue strength, since it relates to the ability of the picture cue to elicit imagery that illuminated elements of the respondent's personality.

2.3.2. Picture cue selection in the McClelland et al. (1953) PSE.

McClelland and colleagues (McClelland et al., 1953) sought to improve the

reliability and validity of the measure when they repurposed the TAT to their goal of implicit motive assessment. Given their more specific objective of arousing and measuring the achievement motive (rather than personality in general), they found many of Morgan and Murray's (1935) TAT images to be unsuitable, given that they rarely gave rise to stories featuring motive-relevant content (i.e., the cue strength of the images was too low). In order to address this problem, McClelland et al. (1949) supplemented two of Morgan and Murray's (1935) more achievement-related TAT pictures, "*father*" talking to "*son*" (picture 7BM) and *boy with vague operation scene in background* (picture 8BM), with two novel images "chosen to suggest achievement—either at a specific task or general level and in school-related and unrelated situations" (McClelland et al., 1949, p. 245)—subsequently referred to as *two men* ("*inventors*") in a shop working at a machine (often abbreviated to *inventors*) and *boy in checked shirt at a desk, an open book in front of him* (often abbreviated to *boy in checked shirt*). Since the objective of McClelland et al.'s (1949) study was to investigate whether the content of imaginative stories reflected states of experimentally induced motive arousal, the confirmation of their hypothesis suggested that the experimental conditions (a combination of the setting, instruction, and picture cue content) were sufficient for the arousal of the achievement motive, and so was taken as proof that the images themselves were suitable. Consequently, these four images became part of McClelland and colleagues' *standard set*: an eight-image battery that would regularly be used in their investigations of the achievement motive. In addition to the images mentioned earlier, the standard set also included another card from the Morgan and Murray (1935) TAT, *young boy with violin* (picture 1), and three other images: *two men in colonial dress, printing in a shop*, "*cub reporter*" scene, and "*lawyer's*" office.

2.3.3. *Picture cue selection in the Heckhausen (1963b) PSE.*

While the McClelland et al. (1953) PSE collapsed the approach and avoidant components of the achievement motive into a single motive score, Heckhausen's modified assessment tool was designed specifically to maintain a distinction between hope of success and fear of failure. Just as his content coding categories were devised in order to exclusively capture facets of either approach- or avoidance-related achievement motivation, his picture selection criteria were driven exclusively by a desire to reflect this theoretical distinction between HS and FF. In order to ensure to sufficient pull for HS and FF across his picture set, Heckhausen chose images that unmistakably depicted either hope of success (a student sitting at a desk and smiling happily) or fear of failure (an anxious-looking office worker hesitating outside the door to his director's office), clearly emphasizing cue strength at the expense of cue ambiguity (Heckhausen, 1963b; cited in Brunstein & Heckhausen, 2008).

2.3.4. *Picture cue selection to the present day.*

As the introduction to this section suggested, the images developed by these pioneering early researchers continue to be widely used in implicit achievement motivation research. In the years following the publication of *The Achievement Motive*, researchers have made widespread and consistent use of images from the standard set when conducting PSE assessment of *nAch* or related constructs (e.g., *boy in checked shirt* has been used by, amongst others, Elliot & Sheldon, 1997; Orso, 1969; Teevan & Stamps, 1973; and Thrash, Elliot, & Schultheiss, 2007), with their usage frequently justified on the grounds of previously demonstrated validity. While some researchers in the McClelland/Atkinson tradition conducted in-depth studies both to confirm the efficacy of

the standard set images and to pre-test new ones (e.g., Birney, 1958; Haber & Alpert, 1958; Jacobs, 1958), the composition of the most widely used picture sets remained largely unchanged, with researchers continuing to rely on intuition and the assurances of previous validation when selecting images for use in PSE research. Birney (1958) represents a particularly notable exception. He did much to systematize (and to some extent legitimize) McClelland et al.'s use of face validity as the sole picture selection criteria, demonstrating that *n*Ach scores derived from a large pre-test battery of 18 images correlated with subjective judgments of achievement-relatedness, suggesting that effective images can initially be identified by subjective appraisal. However the number of pictures in widespread use did not proliferate. More pictures were eventually added to the core battery by McClelland, Atkinson and colleagues, such as *architect at desk*, *trapeze artists* and *two women in lab coats in laboratory* (McClelland, 1975), and this expanded version of the standard set has become a convenient repository for researchers searching for appropriate achievement assessment cues.

Much like the associated coding system, Heckhausen's (1963b) pictures were not widely used outside of German-speaking academia for several decades following the publication of his need achievement measure, with researchers typically opting to use the more well-known McClelland et al. (1953) coding system and associated picture set. That situation changed however with Schultheiss' (2001) publication of his English translation of the Heckhausen measure, which also included reproductions of Heckhausen's original picture cues. This translation made the measure more accessible to the global research community, and the 21st century has seen a limited resurgence in the use of Heckhausen's stimulus images. For example, *director's door* has recently been used in research by Pang,

Villacorta, Chin, and Morrison (2009). Nonetheless, the images from the McClelland tradition remain by far the most widely-used, facilitated in part by the influential publication of Smith's (1992) *Motivation and Personality: Handbook of Thematic Content Analysis*, which included reproductions of some of McClelland and colleagues most frequently utilized images.

While there have been a few recent attempts to develop and validate new images (e.g., *bicycle race*, Stanton & Schultheiss, 2007; *soccer duel*, Schultheiss & Rohde, 2002; *chemist*, Pang 2010b), contemporary researchers rely largely on pictures and picture sets that are several decades old, employing less than rigorous selection criteria, such as the aforementioned “intuition or on personal advice by other scholars” (Schultheiss & Brunstein, 2001, p. 72) while collectively assuming the continued reliability of these images, despite repeated warnings to the contrary (Henry, 1956; Murstein, 1972; Pang 2010b; Smith et al., 1992). The problems and pitfalls associated with the continued usage of these picture cues and sets will be explored further in the next section.

2.4. Picture cue selection in contemporary research: Problems and pitfalls.

The introduction to this thesis hinted at possible issues with the continuing use of older images and the reluctance of researchers to develop novel PSE images. In this section, three of the most serious potential problems associated with this practice will be outlined: (a) the lack of independent validation of picture sets, (b) declining validity due to datedness and bias, and (c) questionable suitability for the assessment of HS and FF.

2.4.1. Lack of independent validation.

While McClelland and colleagues were certainly aware of the need to use images

with sufficient pull for *nAch* (hence their decision to introduce images chosen to suggest achievement in their seminal 1949 study), their image selection criteria were largely unsystematic. In the construction of the standard set they chose images that possessed face validity, meaning that in their own subjective judgment, these images were deemed likely to be effective in their arousal of the achievement motive. As their continuing research program provided evidence for the reliability and validity of the combined coding system/picture set assessment instrument (the nascent PSE), their results were taken as vindication of their initial picture cue selection criteria, and as justification for the continuing use of the standard set. At no point was the validity of their picture set assessed independently of that of their newly developed coding system. Indeed, it would have been impossible for McClelland, Atkinson and colleagues to make an independent assessment during their pioneering work, since they were concurrently developing the very first empirically derived content coding system, and hence there was no alternative yardstick against which the validity of their picture set could be measured. And, so long as the combined instrument demonstrated good psychometric properties, there was little reason to be concerned with the specific contribution of the picture cues to the PSE's validity and reliability. Such an unsystematic approach only becomes problematic when successive generations of PSE researchers assume the continuing suitability of these images despite mounting concern that their validity may have been compromised.

2.4.2. Declining validity due to datedness and bias.

The efficacy of a projective measure like the PSE is contingent on the tendency of respondents to project their internal states onto the stimuli. In order for projection to occur, it is universally acknowledged that respondents must be able to identify with the

depicted actors, viewing them as being generally similar to themselves in accordance with the principle of relevance (or identification in Murray's terminology) outlined in section 2.1.3. As picture cues age, the scenes they depict become progressively more dated, featuring tasks, clothing, technology, and social conventions that have since fallen out of favor or been rendered obsolete. Several examples can be found in McClelland et al.'s widely used standard set. *Trapeze artists* was developed at a time when the circus was a popular form of family entertainment, whereas today very few people will have ever attended a circus, and even fewer would deem it to be a personally relevant recreational activity. Similarly, *inventors* shows two men dressed in old-fashioned aprons, shirts and suspenders fixing a rather archaic-looking machine in a workshop. While such a task may have been typical of many blue-collar occupations in the mid-20th century, it is likely to be of little relevance to contemporary respondents, most of whom will work (or to aspire to work) in office jobs, handling computers and spreadsheets rather than chisels and lathes. Given this disconnect between the lives of respondents and the content of PSE stimuli, it seems highly likely that the principles of relevance and universality have been compromised, leading to a reduction in the effectiveness of these picture cues. Such concerns were voiced by early researchers (e.g., Henry, 1956; Murstein, 1972, cited in Smith et al., 1992) who commented that "pictures that are out of date with respect to such features as clothes and hair styles may not elicit stories that are characteristic of current experiences", and advised that "it may be necessary to update standard pictures from time to time" (p. 518). Despite these warnings, which have been restated by other researchers in recent years (e.g., Pang, 2010b), researchers have continued to make use of pictures from the McClelland/Atkinson standard set, some of which are nearly as old as the

discipline of psychology itself.

Datedness is not the only problem associated with the standard set. The McClelland/Atkinson research tradition was afflicted by a long-standing gender bias (the McClelland et al. coding system was originally derived through analysis of stories written by men in response to pictures of men), and it was not until the 1970s that researchers in this tradition began using picture cues featuring women, such as *two women in lab coats in laboratory* (McClelland, 1975). Despite these later efforts, few of today's commonly used *nAch* pictures feature female actors, with the aforementioned *women in lab* being the only notable example. While evidence of the effects of gender bias is lacking, researchers have found significant interactive effects of participant gender and picture character gender on story content (Worchel, Aaron, & Yates, 1990), while other researchers have noted recent attempts to address gender imbalance (Brunstein & Heckhausen, 2008) or recommended systematic study of gender bias as an important topic for future research (Pang, 2010b).

Similarly, the pictures in the standard set are consistent in their depiction of Caucasian actors, with no other ethnic group being clearly represented. Such an ethnic imbalance suggests that these picture cues may be less relevant to non-Western, non-white individuals, rendering the PSE less sensitive to implicit motives in these cultures. Evidence for this assertion is provided by Hofer (2010), who suggests that insensitivity to cultural nuances might lead to item bias: a source of systematic error in measurement that arises when similarly motivated individuals from different cultures react differently to a given picture cue. The presence of such bias decreases the cross-cultural validity of the measurement instrument, since individual motive scores and sample means can no longer

be meaningfully compared across cultures. Hofer and colleagues have documented such bias in the PSE responses of German and Zambian adolescents (Hofer & Chasiotis, 2004) and German, Cameroonian, and Costa Rican adults (Hofer, Chasiotis, Friedlmeier, Busch, & Campos, 2005), and have identified four unbiased picture cues for assessment of implicit power and affiliation motivation in culturally diverse samples (Hofer et al., 2005).

2.4.3. Suitability for the assessment of HS and FF.

Yet another problem with the standard set pertains to its suitability for the assessment of HS and FF. While it is well established that acknowledging the fundamental distinction between the approach and avoidance subcomponents of achievement motivation increases the scope of achievement motive research (Schultheiss, 2001), it is currently unknown whether the standard set's constituent images possess sufficient cue strength or cue ambiguity for the valid assessment of HS and FF. Currently, researchers utilizing pictures from the standard set or from the wider McClelland/Atkinson tradition in their assessment of HS and FF (e.g., Engeser & Rheinberg, 2008; Langens & Mörth, 2003; Thrash, Elliot, & Schultheiss, 2007) assume the validity of these images in the absence of supporting empirical evidence. Clearly, a picture set developed and validated specifically for the assessment of HS and FF would offer a superior alternative to the standard set in such instances. However, while contemporary researchers in the Heckhausen tradition do occasionally make use of the images originally used by Heckhausen in the development of his revised need achievement measure (see section 3.7), use of these picture cues is equally problematic. Firstly, these images have been employed only infrequently, and thus don't possess the historical legacy that is used to justify the continuing use of the standard set. Secondly, the issues of datedness and

gender/ethnic bias that afflict the McClelland/Atkinson standard set also affect the Heckhausen picture set. Of the six images, none feature either a female or a non-Caucasian protagonist, while all could be considered dated given their depiction of atypical professions (pictures C and F), old-fashioned attire (pictures A and E), or obsolete technology (pictures D and E). As such, there is reason to suspect that the principles of relevance and universality will have also been compromised in the Heckhausen picture set, potentially leading to a decline in cue strength and validity. Taken together, these observations illustrate that there is currently no widely used, adequately validated picture set for the assessment of hope of success and fear of failure, a significant shortcoming that must be addressed if research in the field is to flourish.

2.5. Recent attempts to standardize and improve picture cue selection.

In the past decade, several attempts have been made to address some of the concerns listed in section 2.4. For the most part however, these studies have focused on better characterizing existing picture cues in terms of their ability to elicit specific types of motive imagery, the rationale being that a more comprehensive understanding of their motivationally arousing properties will allow for better cue selection in PSE research. Schultheiss and Brunstein (2001) provided cue strength data for six images used in the McClelland/Atkinson tradition—*architect at desk*, *women in laboratory*, *ship captain*, *couple by river*, *trapeze artists*, and *nightclub scene*—demonstrating significant between-picture differences in the tendency to arouse *nAch*, *nAff*, and *nPow* coded using the Winter (1994) system. The authors went on to identify images that pulled effectively for each of the motives, noting that scores derived from moderate-high pull images typically

exhibited enhanced variance, which suggests these images may be more sensitive to dispositional differences in underlying motives. Schultheiss and Brunstein reasoned that by attending to these results, PSE researchers should be better placed to discern between effective and ineffective picture cues when planning their research. Langan-Fox and Grant (2006) noted the absence of a “standard battery of cues for assessing the big three motives”, citing Wyatt’s (1947) claim that “the best prospect for practical advancement in the use of the test (TAT)...seems to lie...in the uniformity of procedures and of the picture stimuli” (p. 279). In order to address this methodological shortcoming, the authors adopted a similar approach to Schultheiss and Brunstein (2001), selecting a pool of widely used images from the McClelland/Atkinson research tradition and comparing their ability to elicit achievement-, affiliation-, and power-related content in the PSE. Unsurprisingly given the similarity of their approach and choice of picture cues, the results of Langan-Fox and Grant (2006) echoed those of Schultheiss and Brunstein (2001), demonstrating significant between picture differences and identifying high-pull images that they suggested may be better suited for the assessment of a specific motive (e.g., *women in lab* for *nAch*).

While the value of such work is undisputed, it is clear that much work remains to be done if PSE picture cue selection is to be significantly improved. Documenting cue strength is not the same as demonstrating validity, and so it is premature to claim that a picture set provides a better assessment of implicit motives when the validity of the resulting scores is not critically examined through comparison with other supposedly less suitable images. Only if these picture sets give rise to scores that better predict theoretically relevant criteria can it truly be said that they represent an improvement over

other widely used picture cue batteries. Furthermore, it is unfortunate that few researchers have attempted to develop and validate an entirely new picture set with a view to enhancing relevance and universality (e.g., by featuring contemporary scenarios, addressing issues of gender and ethnic bias), an approach that should theoretically yield significant improvements in predictive validity. Finally, it should also be noted that few attempts have been made to ascertain whether these commonly used images are suitable for the assessment of HS and FF, or to develop new picture sets that better reflect this important distinction.

Pang (2010b) took an important first step towards addressing these shortcomings by developing a new picture set for the assessment of *nAch*. Importantly, this picture set includes novel images alongside those used in prior research. Beginning with a pre-test battery of 11 images (two from the McClelland/Atkinson tradition, two from Heckhausen's original 1963 picture set, and seven new images culled from a variety of print and media sources), she compared the cue strength of these images across the achievement, affiliation and power domains, arriving at a new six-picture set comprising one previously validated image (*women in lab*) and five novel images (*skaters*, *piano lesson*, *soccer duel*, *chemist*, and *gymnast*) that exhibited moderate-high pull for *nAch* and weak pull for alternate motives (i.e., the images possessed relatively low cue ambiguity). This new *nAch* picture set represents a valuable first attempt to enhance relevance and universality by utilizing more modern picture cues featuring multiple depictions of female characters. However, validation data for this picture set is somewhat limited (Pang et al., 2009) and to date, no attempt has been made to compare the effectiveness of this novel picture with older, more commonly used picture sets.

2.6. The present research.

The present research attempts to address each of the concerns listed previously by developing the first entirely new picture set for the assessment of implicit achievement motivation: a picture set that offers significant, empirically verified psychometric improvements over commonly used picture sets. Furthermore, it provides the first empirical investigation of set ambiguity, with a view both to improving general picture selection guidelines and enhancing the validity and reliability of my new picture set. In keeping with the weaknesses highlighted in the preceding sections, the improvements offered by this new picture set are fivefold:

- a) Improved relevance and decreased bias: The new picture set features novel images specifically chosen to address issues of datedness and ethnic/gender bias. These selection criteria enhance the relevance of the new pictures, and improve the validity of the new picture set as a whole. This proposition was empirically assessed by comparing the new picture set with a comparison picture set as described earlier.
- b) Suitability for the assessment of HS and FF: The suitability of the new picture set for the assessment of HS and FF was assessed specifically. Having demonstrated good psychometric properties, the new picture set offers an improved alternative for those researchers currently investigating HS and FF using picture sets that have been validated only in terms of *nAch*.
- c) Independent validation: The reliability and validity of this novel picture set were assessed in conjunction with the previously validated Heckhausen (1963b) need achievement measure, constituting an important first attempt to validate a picture

set independently of its associated coding system. Whereas the validation data for the McClelland/Atkinson and Heckhausen images reflected the combined validity of the coding system-picture set instrument, the present study examined the validity of the picture set in isolation, meaning that any psychometric improvements can be attributed entirely to the picture cues.

- d) **Systematic comparison:** The reliability and validity of the new picture set were systematically compared to that of a comparison picture set comprising commonly-used images from the both the McClelland/Atkinson and Heckhausen traditions. Rather than comparing image suitability by using a proxy variable such as cue strength, the present design allows for direct comparison on measures of validity and reliability, meaning that the observed improvements are indicative of the enhanced suitability of the new picture set.
- e) **Appropriate set ambiguity:** Given the absence of empirical data on the role of set ambiguity, the present research compared the psychometric properties of two new picture set versions: one ambiguous and one unambiguous. The better performing picture set was chosen, offering further improvements in validity and reliability.

3. The Content Coding System

The development of systematic content coding systems transformed the TAT from an exploratory psychoanalytic tool to a more scientific measure of personality. Without these systems, the conclusions drawn from analyzing TAT stories languish in subjectivity, and cannot attain the levels of reliability and validity required of a psychological measure. Over the course of the PSE's development, the evolution of these content coding systems has been driven by both data and theory, leading to a diversification in the systems available to PSE researchers. This proliferation has been both a blessing and a curse, enhancing our understanding of the achievement motive yet causing difficulties in integrating the findings of different methodological traditions.

3.1. Content coding in the TAT.

In the PSE's predecessor the TAT, motive assessment was not standardized. Morgan and Murray (1935) originally intended that the narrative output generated by the TAT would be "interpreted rather than scored" (Winter, 1999, p. 107), with clinicians encouraged to rely on their own experience and judgment when inferring the presence of psychological constructs or states. Nonetheless, it would be inaccurate to claim that the TAT had no scoring system at all. In his famous study *Explorations in Personality*, Murray (1938) outlined his structural theory of *personology*, which identified 29 fundamental needs, 20 presses (environmental forces regulating the expression of needs) and their various interactions. This taxonomy of personality gave rise to Murray's elaborate need-press-thema scoring system, which was published along with the accompanying picture cards as a revised version of the TAT (Murray, 1943). Scores on the

various aspects of the need-press-thema system were quantifiable, yet were not critical in the diagnostic procedure, and TAT administrators were free to bring their own theories to bear on the final result (Gieser & Stein, 1999). As the TAT grew in popularity, many psychologists adopted Murray's need-press-thema system (e.g., Combs, 1946; Sanford, Adkins, Miller, & Cobb, 1943), with some adapting and modifying this system through inclusion of new scoring categories (e.g., Aron, 1949; 1950). Some researchers and clinicians went further, devising new systems that varied greatly in structure, complexity, emphasis on quantitative or qualitative analysis, and in the adoption of an idiographic or nomothetic approach (Winter, 1999). As the number of ways of interpreting and scoring the TAT proliferated, standardization became progressively more unlikely, providing yet more ammunition for critics such as Eysenck (1968) who claimed that measures relying on subjective inference fail to meet the psychometric standards of reliability and validity, and therefore should not be referred to as tests at all.

3.2. Content coding in the PSE.

These psychometric deficiencies were soon to be addressed by a group of researchers led by two influential figures: David McClelland and John Atkinson. Schooled in the staunchly empirical behaviorist tradition, McClelland had little to do with the emerging field of personality psychology until circumstance dictated he teach a personality course in the years immediately following the Second World War. During this time he supervised the honors thesis of John Atkinson, who proposed using projective measures to develop psychometrically sound measures of psychogenic needs. Atkinson (1947) noted that "experimental demonstration of this underlying assumption has not yet

been accomplished” (p. 55), and proposed a series of experiments investigating the effects of hunger—a simpler and more easily operationalized need—on perception. Initially, the two researchers did not utilize the TAT. Instead, they adapted Miller's (1939) subliminal perception method, depriving university students of food for either two or 16 hours, before asking them to identify various images at low levels of illumination, the premise being that food-deprived individuals (i.e., those in a state of hunger motive arousal) would be more likely to identify the obscured objects as foodstuffs or items associated with food preparation. This hypothesis was borne out by the data, convincing McClelland and Atkinson of the utility of projective measures in the study of motivation.

The two researchers went on to replicate this initial finding in a study of submariners (McClelland & Atkinson, 1948). However, their observation of relatively weak (albeit still significant) effects encouraged them to experiment with alternative methods, leading McClelland and Atkinson to try the TAT, with which both of them had a passing familiarity. In the same study, McClelland and Atkinson deprived individuals of food for one, four, or 16 hours, then asked them to write imaginative stories in response to the various TAT cards. The results were striking. Food-deprived individuals wrote stories in which food was a more central theme, while need for food and relevant instrumental activity were also mentioned more frequently by members of the hunger motive arousal group. Further convinced that projective measures were the key to measuring motivation, the researchers turned their attention to Atkinson's initial objective: the measurement of more complex psychogenic needs, such as the need for achievement.

3.3. Measuring *nAch*: The McClelland et al. (1953) system.

In their seminal study, McClelland, Clark, Roby and Atkinson (1949) had two primary objectives. Firstly, they sought to investigate whether more psychologically elaborate motives exerted the same effects on projection as physiological needs such as hunger. If this was the case, then projective measures such as the TAT could be used to assess higher-order motivational constructs, such as the need for achievement. Secondly, they hoped to solve one of the TAT's perennial problems: the lack of a standardized, objective coding system for interpreting generated narratives. McClelland and his colleagues were acutely aware of this shortcoming, noting that "one of the crucial problems in this type of experiment is to find a scoring system for thematic stories that is objective enough to provide high observer agreement and sensitive enough to reflect changes in motivational states" (p. 243). They hoped that by developing such a coding system themselves, subsequent research on the achievement motive could be conducted in a more rigorous and scientific manner, free from the subjective uncertainty of traditional TAT assessment.

3.3.1. Development.

The key to McClelland et al.'s (1949) contribution was their decision to differentially arouse the achievement motive in their participants. Building on their previous work, in which the hunger motive was aroused to greater or lesser extents through food deprivation, McClelland and colleagues experimentally manipulated the achievement motive by varying the instructions accompanying various achievement tasks. If the modified TAT was suitable for the assessment of such a need, they reasoned, then these different states of arousal should be reflected in the content of their participants'

imaginative stories. Furthermore, a content coding system that is constructed in order to capture the fundamental differences between narratives generated in high and low arousal conditions should provide the most valid assessment of this latent construct, facilitating future investigation by providing a demonstrably effective coding solution.

To this end, McClelland et al. (1949) gave simple pencil and paper tests (such as anagrams, scrambled words, and motor perseveration tasks) ostensibly measuring intelligence and psychomotor skills, to groups of male college students, many of whom were ex-servicemen, manipulating the arousal of the achievement motive by varying task descriptions across six conditions: neutral, relaxed, achievement-oriented, success, failure, and success-failure. By comparing the content of stories written under the success, success-failure and relaxed conditions, McClelland and colleagues observed that protocols generated under conditions of motive arousal contained less unrelated and task achievement imagery, and more “general achievement imagery, achievement-related deprivation themas, stated needs, successful instrumental acts, anticipatory goal responses, nurturant or hostile press, and positive affective states” (p. 254), categorical differences that would subsequently form the basis of the first empirically derived content coding system for measuring achievement motivation (McClelland, et al., 1953).

The McClelland et al. system is considered to be empirically derived because the categorical differences emerged through contrasting data generated by participants in differing states of experimentally induced motive arousal, rather than being decided upon in an *a priori* fashion. Nonetheless, the conception of the system was not an entirely bottom-up process, and McClelland et al. drew upon two main theoretical perspectives when codifying these emergent differences in the data. Firstly, they borrowed from

Murray's (1943) TAT scoring system by including need, press, and thema categories with very similar definitions. Secondly, they derived further content coding categories through reference to McClelland's previous work in the behaviorist tradition (McClelland, 1942), in which the sequence of motivated behavior is broken down into several parts, an instigation (i.e., a need), instrumental acts, and goal responses. As such, the McClelland et al. system should be considered to be both theoretically and empirically derived.

3.3.2. Structure and specification.

The McClelland et al. coding system consists of 15 separate categories. In the first instance, raters must decide whether the story features achievement imagery (AI), defined by the authors as story characters aiming for "success in competition with some standard of excellence" (p.110). According to McClelland and colleagues, this striving usually manifests in one of three forms: competition with a standard of excellence (e.g., hoping to win an essay contest), unique accomplishment (e.g., discovering the cure for a deadly disease), or long-term involvement in the pursuit of an achievement goal (e.g., building a successful business over many years). If such imagery can be identified, the story is classified as achievement-related, and AI is scored with a value of one mark. Only if AI is scored can the rater proceed to code the various subcategories described below. Stories that are not scored for AI are classified in terms of whether they contain doubtful imagery (TI) or unrelated imagery (UI), the presence of which precludes further scoring at the subcategory level. Stories characterized by TI typically make reference to achievement without fulfilling any of the criteria for scoring AI (e.g., by describing engagement in an achievement-related task without a standard of excellence). Scoring TI results in a score

of zero for that story. UI on the hand refers to imagery that is entirely unrelated to achievement. Stories characterized by UI may feature imagery relevant to an alternate motive, such as *nAff* or *nPow*, or may feature mundane imagery lacking in motivational content. Scoring UI leads to a deduction of one mark from the participant's total.

Once the presence of AI is established, referral to a number of internal states or overt actions relevant to goal pursuit can be scored as specific elaborations of achievement imagery. Each subcategory can only be scored once per story (although see additional restrictions for instrumental activity categories), and the scoring of each subcategory is worth one additional mark when calculating the total for that picture. All subcategories, positioned in terms of their hypothesized role in the adjustive behavioral sequence (McClelland, 1942) can be found in figure 1.

Figure 1. Position of McClelland et al.'s (1953) categories in the adjustive behavioral sequence

The first subcategory, need (N), is scored whenever someone in the story states a desire to reach an achievement goal, or has such a desire attributed to them by the author. Anticipated success (Ga+) is scored whenever a story character anticipates successful attainment of an achievement goal, while anticipated failure (Ga-) is scored whenever a story character expects to be unsuccessful. During the course of this activity, story characters may experience positive affect arising from goal attainment (G+) or negative affect arising from failure (G-), both of which are scored as instances of achievement imagery. Furthermore, story characters may execute instrumental activities that successfully result in goal attainment (I+), or they may engage in instrumental, goal-directed activities that are unsuccessful (I-) or of uncertain outcome (I?). These actions are also scored for achievement imagery, although since (I) can only be scored once per story, it is unclear what benefits are derived from this distinction. Unsuccessful activities may be blocked by obstacles either residing within the person (Bp) or by external forces beyond the individual's control (Bw), while successful actions may be accompanied by nurturant press (Nup), defined as help or sympathy that aids a character in their goal-directed behavior. Such instances of help or hindrance are scored. Finally, a story is scored again for the presence of an achievement thema (AchTh), which occurs when achievement imagery is "elaborated in such a manner that it becomes the central plot or thema of the story" (McClelland et al., 1953, p. 136). Summing across the various categories yields a total achievement score with a maximum value of 11 and a minimum value of negative one. Typically, these picture-specific values are then summed or averaged across the entire picture set to give a single score representative of the underlying motive.

3.3.3. *Validation.*

The validity of McClelland et al.'s system in measuring implicit achievement motivation has been confirmed in numerous studies, both behavioral and physiological. In the years immediately following McClelland et al.'s (1949) ground-breaking study, this core group of researchers gathered much data regarding the behavioral correlates of *nAch*. Atkinson (1950) found that under conditions of achievement motive arousal, individuals dispositionally high in *nAch* completed more tasks than those lower in *nAch*, while Lowell (1952) found that high *nAch* individuals became increasingly more productive as a scrambled word task progressed, in contrast to low *nAch* individuals whose productivity peaked then went into decline. Compared to those low in *nAch*, achievement-oriented individuals also perform better in mental arithmetic (Wendt, 1955) and in a six-digit substitution task (French, 1955). These results provided early indications that achievement-motivated individuals performed better and more persistently on intrinsically motivating tasks under conditions of appropriate motive arousal. Furthermore, in an influential series of studies assessing *nAch* and risk preference, Atkinson (1957) found that high *nAch* individuals (those with a high resultant achievement motivation, for a more detailed discussion of this somewhat problematic conceptualization see section 3.5.2.) prefer to work on moderately difficult tasks, a highly robust finding that can be understood in terms of his expectancy-value theory of motivation, which states that task preference (i.e., the strength of the motivation to performance a specific task) is a multiplicative function of dispositional motivation, expectancy of success, and the incentive value of goal attainment. Since Atkinson assumes that the incentive value of success is a positive linear function of difficulty (i.e., very difficult tasks carry high

incentive value, while very easy tasks have a low incentive value), tasks of moderate difficulty—those that are moderately incentivizing and have a moderate subjective probability of success—maximize the value of this multiplicative function and thus lead to enhanced preference. This preference has been documented in numerous studies (e.g., Atkinson & Litwin, 1960; de Charms & Carpenter, 1968).

Physiological evidence also speaks to validity of McClelland et al.'s system. When compared to those scoring low in *nAch*, high scorers have been shown to exhibit a higher level of physiological activation when engaged in achievement-related activities. Mücher and Heckhausen (1962) observed increased muscle tension in higher scorers compared to low scorers when concentrating on a task, while Raphelson (1957) found high *nAch* scorers consistently exhibit a higher galvanic skin response, which is indicative of greater activation of the sympathetic nervous system. Wendt (1955) found that those high in *nAch* exhibit an increased critical flicker fusion frequency compared to low *nAch* scorers, indicating a greater awareness and vigilance when concentrating on a task, while McClelland and Lieberman (1949) demonstrated that high *nAch* scorers perceive achievement-related words more quickly when presented for split-second intervals using a tachistoscope.

Research during the latter part of the 20th century confirmed and elaborated upon these initial findings, painting a consistent picture of the achievement motive and comprehensively validating McClelland et al.'s need achievement measure. Generally, McClelland et al.'s *nAch* measure has been found to predict operant achievement-related behaviors over time. Achievement motivated individuals regularly participate in more challenging leisure activities (Veroff, 1982), get promoted faster (McClelland & Boyatzis,

1982), and are more inclined towards entrepreneurship (Collin, Hanges, & Locke, 2004). Importantly, Spangler (1992) conducted a meta-analysis of achievement motive research (the vast majority of which utilized the McClelland et al. *nAch* coding system), finding that in the presence of intrinsic, or task-related, achievement incentives, *nAch* strongly predicts achievement-related outcomes and behavior.

3.3.4. Weaknesses.

Despite its widespread use in the latter half of the 20th century, there are several major limitations of the McClelland et al. system. First and foremost, the system fails to capture the important distinction between HS and FF. Although McClelland and colleagues were aware of this distinction, they persisted with an instrument that obscures and ignores the fundamental distinction between these two submotives. This significant shortcoming is discussed in greater detail in section 3.10. Secondly, the definition of achievement imagery in the McClelland et al. system is too broad, encompassing aspects of both power motivation (competition in non-achievement settings) and affiliation motivation (altruistic assistance). Consequently, the McClelland et al. system is not a pure-bred measure of achievement motivation (Schultheiss & Brunstein, 2005), since it is contaminated with partial measurement of other implicit motives. Thirdly, the McClelland et al. system is complex and time-consuming to administer, a weakness borne of the need to first classify the story as either AI, TI, or UI before proceeding to score the various subcategories. This issue is compounded by occasionally confusing training materials that do not make explicit the links between scoring and particular narrative elements, hampering the ability of novice scorers establish proficiency. It is also important to note that the vast majority of evidence supporting the validity of the McClelland et al. system

has been gathered in studies of Western, undergraduate samples. This limitation is shared by each of the four major assessment tools, all of which have been used only infrequently outside of Europe and North America.

3.4. Measuring *n*Ach: The Winter (1994) system.

In the development of his *Manual for Scoring Motive Imagery in Running Text*, David Winter (1989; 1991; 1994) sought to address some of the disadvantages mentioned earlier. Specifically, he wanted to create a simplified (yet equally valid) content coding system that allowed simultaneous assessment of the big three motives and could be applied to a more diverse range of source materials (Winter, 2010). Whereas the McClelland et al. system required discrete, independent narratives—the golden rule being that each category cannot be scored more than once per story—the Winter (1994) system dispensed with this requirement in order to facilitate the analysis of longer, more complex texts, such as speeches, correspondence, interviews, and works of fiction. Despite this original intention to liberate motive content analysis from the confines of the PSE, the efficiency and comparative simplicity of the Winter system also appealed to researchers in the PSE tradition, and it eventually superseded the McClelland et al. system as the most commonly used PSE measure of implicit achievement motivation (Weinberger, Cotler, & Fishman, 2010).

3.4.1. Development.

The Winter (1994) system was developed by combining and simplifying the three most prominent PSE measures of the big three motives: the McClelland et al. (1953) system for measuring achievement motivation, the McAdams (1980) system for scoring affiliation/intimacy motivation, and the Winter (1973) system for scoring power

motivation. Previously, researchers who wished to apply these systems to non-PSE materials were forced to make *ad hoc* modifications in order to make scoring possible (e.g., Donley, 1968; Donley & Winter, 1970; Winter & Stewart, 1977). Winter's approach was to codify these various modifications into a single, all-purpose system that could be applied to texts of any length, even those that are not naturally subdivided into separate segments. Berlew (1956) circumvented the McClelland et al. (1953) system's restrictions by dropping all achievement subcategories altogether and simply scoring achievement imagery per hundred lines of text, while Donley and Winter (1970) built on this innovation by expressing scores in terms of motive imagery per thousand words, a modification that has been widely adopted as a way of controlling for verbosity or linguistic ability. Winter and Stewart (1977) simplified several coding systems in order to improve their application to political speeches, subtly altering the definitions of motive imagery for the sake of simplicity and clarity. These modifications were consolidated to form Winter's (1991) *Manual for Scoring Motive Imagery in Running Text*, which was subsequently revised (Winter, 1994). Scores from the Winter system were found to be highly correlated with scores derived from the original systems (Winter, 2010), and could successfully discriminate motive-aroused individuals from non-aroused individuals in the same manner as their progenitors. Each motive score also demonstrated good test-retest reliability (r_s .46 to .60; Winter, 2010), and predictive validity (further discussed in section 3.4.3).

3.4.2. Structure and specification.

The simplicity of the Winter *nAch* subsystem is largely due to his abandonment of McClelland et al.'s system of subcategories. Rather than first identifying the presence of

achievement imagery before proceeding to code for different subtypes, the Winter system simply involves scoring any “action...wish, concern or other internal state” (Winter, 1994, p. 4) relating to the pursuit of “a standard of excellence” (Winter, 1994, p. 8) as an instance of achievement imagery. According to Winter, such standards are usually expressed in one of five forms: (a) adjectives that positively evaluate performance; (b) goals or performances that are described in such a way as to suggest positive evaluation; (c) mention of winning or competing with others; (d) failure, doing badly or lack of excellence; or (e) a unique accomplishment. These various instances were considered to capture the most important elements of McClelland et al.'s original system.

Unlike the McClelland et al. system, the Winter system sets no upper limit on how often achievement imagery can be coded in the course of a single piece of text. There are, however, regulations governing when and how achievement imagery can be coded. The sentence is the unit of analysis, and achievement imagery may only be scored once per sentence. A further restriction is that achievement imagery cannot be scored in consecutive sentences, unless these two occurrences are separated by an instance of alternate-motive imagery (i.e., *nAff* or *nPow*). These rules are intended to discriminate between instances of repetition or clarification, and those that truly represent a preoccupation with achievement. When the Winter system is applied to PSE research, individual scores are typically calculated in the same way they would be in the McClelland et al. system, by average or summing picture-specific scores across the entire battery.

3.4.3. Validation.

The bulk of the early work validating the Winter running text system was

conducted by David Winter himself in an extensive program of research concerning the motivation of historic political leaders. Using his newly developed system, Winter analyzed the content of speeches, interviews and governmental documents produced by various U.S. presidents, documenting significant associations between their implicit motivation and different outcome measures, such as ratings of greatness, incidence of scandal, and the number of wars entered (for a summary, see Winter 2002; 2005). While a number of intriguing significant associations were found for power motivation (positively related to war entry and perceived greatness), achievement motivation was found only to relate to historians' ratings of idealism.

Fortunately, researchers in other traditions have gathered plentiful evidence supporting the validity of Winter's achievement measure. Some have built on the legacy of McClelland and colleagues by correlating Winter *nAch* scores with indices of entrepreneurial success and business performance. For example, Chusmir and Azevedo (1992) found that the level of *nAch* motive content in a corporation's annual report correlated significantly with its financial return on sales, return on equity, and future growth in sales volume, while Tran and Philipp (2010) documented a significant association between the achievement-related content of tourism organizations' vision statements and spending by tourists. However, perhaps the largest and most compelling body of work supporting the validity of the Winter system's *nAch* subscale comes from the motive congruence literature, where congruence between implicit achievement motivation (as measured by the running text system) and self-reported achievement goals or values have been found to associate with a host of predictor and outcome measures. Brunstein, Schultheiss, and Grässman (1998) found that progress toward motive-

congruent goals, in contrast to progress toward motive-incongruent goals, accounted for students' daily experiences of emotional well-being, while Job and Brandstätter (2009) observed that individuals who focus on achievement-specific affective incentives in their goal fantasies set goals in line with their implicit achievement motivation. More recently, Schüler, Brandstätter, and Sheldon (2012) found implicit achievement motivation to be a significant moderator of the positive effects of competence need satisfaction on both flow and well-being.

3.4.4. Weaknesses.

The Winter system addressed one of the major shortcomings of the McClelland et al. system—its complex and unwieldy application—while empowering researchers to assess multiple motives simultaneously and without the need to master several independent systems. These are significant benefits that have made the Winter system the go-to methodology for implicit motive researchers. However, usability and multimotive scope notwithstanding, the Winter system suffers from the same fundamental drawback as the McClelland et al. system with respect to coding achievement motivation: the conflation of two separate and fundamentally distinct submotives into a single score. This weakness, and the various attempts to address it, will be discussed at length in the remainder of this section.

3.5. *n*Ach, HS, and FF in the McClelland/Atkinson tradition.

Despite the emphasis placed on a unitary achievement motive in their research, McClelland, Atkinson and colleagues were acutely aware that an individual's need for achievement could be further characterized as either hopeful or fearful in nature, a

conceptual distinction that was referenced in some of their earliest works (McClelland & Lieberman, 1949; McClelland et al., 1953). McClelland et al. (1953) contrasted “an achievement motive which is characterized chiefly by a hope of success with one which is characterized chiefly by a fear of failure” (p. 214), and invoked developmental explanations for this apparent specialization, speculating that HS-motivated individuals would have been praised and rewarded for the successful completion of achievement tasks early in life, while FF-motivated individuals would have been primarily punished for their childhood failures. They further surmised that those motivated by HS would have encountered success more frequently than those characterized by FF, who experienced relatively more failure. Reiterating the importance of this distinction, they acknowledged that their measure’s *nAch* score was “a composite which will have to be broken down into various sub-scores” (p. 215); a puzzling statement given their continued use of the unidimensional *nAch* coding system over the ensuing decades. The McClelland et al. (1953) system was never revised by the original authors, and they persisted in measuring a unitary *nAch* construct despite acknowledging its fundamentally dichotomous nature. Indeed, later attempts to incorporate the FF construct into their research merely added to the confusion, as they introduced a self-report measure of FF (described in more detail in section 3.5.2) that was typically administered alongside the PSE measure of *nAch*. It remains unclear why they felt compelled to assess FF separately when they were aware that their projective measure already tapped this construct. The changing conceptualization of HS and FF in the McClelland/Atkinson tradition will be discussed in the following sections.

3.5.1. *The early days: A tripartite division.*

One of the earliest attempts to characterize the HS/FF distinction was made by McClelland and Lieberman (1949), who attempted to assign the two motive tendencies to different sections of the *nAch* score distribution. They divided their participants into three groups (low, medium, and high achievement motivation) on the basis of their *nAch* scores, and attempted to relate these categorical distinctions to differences in their ability to recognize success- and failure-related words that were presented for a very short period of time. High *nAch* participants were quicker to identify success words while medium *nAch* participants were slower to recognize failure words. Low *nAch* participants showed no difference in their ability recognize success versus failure-related words. The authors interpreted these results as evidence that moderately *nAch*-motivated individuals were primarily motivated by the desire to avoid failure. In a similar study, Atkinson (1950) found that individuals in the middle third of the *nAch* score distribution recalled fewer incomplete tasks under achievement motive arousal conditions (compared to neutral conditions), whereas high *nAch* individuals recalled more unfinished tasks when the achievement motive was aroused. In these two studies, moderately *nAch*-motivated individuals demonstrated a deficiency in perceiving or recalling failure-related stimuli. This finding was interpreted as a defensive reaction by McClelland et al. (1953), who suggested that such individuals unconsciously suppressed failure-related cognition in order to avoid the anxiety associated with failure. In line with these results, McClelland and colleagues further suggested that the *nAch* score distribution exhibited a tripartite structure, with the bottom third unmotivated by achievement, the middle third motivated by fear of failure, and the top third motivated by hope of success.

Despite these early results, the notion that the *nAch* distribution can be compartmentalized into a tripartite structure has now been largely abandoned (Heckhausen, Schmalt, & Schneider, 1985, p. 19). Results to be discussed at length in the following sections are not consistent with notion that all FF-focused individuals are moderately motivated, or with the assertion that all moderately *nAch*-motivated individuals are driven predominantly by fear of failure. Even the authors themselves expressed concern that the association of a specific kind of achievement motivation with different levels of the score distribution may be an artifact of the measurement methodology (McClelland et al., 1953, p. 273), and research on the tripartite division slowly ground to a halt as a very different perspective on HS and FF emerged: that of resultant achievement motivation (RAM).

3.5.2. Towards a separation of motives: TAQ and RAM.

The results of these early experiments (McClelland & Lieberman, 1949; McClelland et al., 1953) cemented the notion that anxiety provided the emotional basis for fear of failure, and the view that anxiety underpins avoidant motivation remains popular to this day (Higgins, 1997). However, despite their earlier acknowledgment that the McClelland et al. (1953) coding system assessed FF-related constructs, researchers in the McClelland/Atkinson tradition sought to develop new measures of FF in order to further characterize the distinctions between success-motivated (or achievement-motivated, the two terms were often used interchangeably) individuals and those who were more fearful of failure. Their reasons for doing so remain slightly obtuse, although it may have stemmed from a continuing belief in the unitary nature of McClelland et al.'s *nAch* construct, which would necessitate the development of a separate measure of FF.

In an influential paper, Atkinson and Litwin (1960) introduced the Mandler-Sarason Test Anxiety Questionnaire (TAQ; Sarason & Mandler, 1952) as a novel measure of FF. Citing the TAQ's ability to predict performance decrements in a manner consistent with their notion of fear of failure, as well as previously-documented negative correlations between test anxiety and PSE-assessed *nAch* (Raphelson, 1957), Atkinson and Litwin proposed that the McClelland et al. system and the TAQ provided separate measures of *nAch* and FF respectively, and that FF was fundamentally inhibitory with respect to *nAch*, two ideas that would dominate the thinking of American *nAch* researchers for years to come. In the same article, Atkinson and Litwin documented a preference for moderate risk in high *nAch*/low TAQ participants, and a heightened preference for extremities of risk in low *nAch*/high TAQ participants. This observation that individuals high in *resultant achievement motivation* (RAM; the net effect of *nAch* after subtraction of FF's deleterious effects, assumed to be equivalent to HS) preferred moderate risk while those low in RAM (i.e., higher in FF) preferred either high or low risk was consistent with the predictions made by Atkinson's expectancy-value theory, and so were taken as evidence of the construct validity of the RAM/TAQ method of HS/FF assessment. While some researchers (e.g., Feather, 1962) found further evidence in support of Atkinson's expectancy-value theory using the RAM/TAQ methodology, others obtained contradictory results. Using PSE-derived measures of HS and FF (described in more detail in the following section), de Charms and Davé (1965) found that both high HS and high FF individuals avoided moderate risk. Such contradictions were suggestive of a disconnect between TAQ and projective measures of FF, and methodological concerns with the TAQ/RAM measure arose as researchers began to question its underlying assumptions

(Heckhausen, 1963b).

Pang (2010a) provides a cogent account of the RAM/TAQ measure's many shortcomings. Firstly, the TAQ confounds self-perceived ability and generalized negative emotionality with motivation to avoid failure. Secondly, the use of a difference score to measure HS illustrates a fundamental assumption that FF is universally debilitating with respect to achievement motivation, an assumption that has been called into question by researchers in the Heckhausen tradition (see section 3.7). Thirdly and most importantly, the fundamental distinction between implicit and explicit motives described in section 1.1 (McClelland, Koestner, & Weinberger, 1989) means that combining explicit measures of FF, such as a self-reported test anxiety, with implicit measures of achievement motivation is inappropriate, given that the two are independent constructs predicting different classes of behaviors. Given these concerns, it seems clear that the TAQ is "neither an adequate nor a valid measure" of FF (Pang, 2010a, p. 47), further highlighting the need for a projective measure of achievement motivation that placed the HS/FF distinction at its core.

3.6. Early projective measures of FF.

Some early researchers were skeptical as to whether fearful, avoidant motives could be assessed by the PSE at all. Scott (1956) hypothesized that the desire to avoid failure-related anxiety might lead FF-motivated individuals to disengage from the PSE's creative writing task if the stimuli were found to be sufficiently threatening, leading to an underestimation of motivation and a sharp decline in validity. In his study, he manipulated PSE task descriptions in a manner similar to McClelland et al. (1949), such that the task

became more ego-involved, and found that avoiders (those individuals who responded with more achievement-related imagery in the neutral condition than in the motive arousal condition) told shorter stories lacking in realism and problem-oriented thinking. Scott's results were far from clear-cut however, and another study in the same paper failed to replicate these effects, casting doubt on the credibility of his conclusions. Other researchers were more optimistic, and attempted to assess FF by considering only the failure-related or deprivation imagery categories of McClelland et al.'s content coding system, a logical choice that would foreshadow Heckhausen's (1963b) development of the first true PSE measure of HS/FF. Clark, Teevan, and Ricciuti (1956) summed scores on the positive and negative categories of the McClelland et al. system to derive measures of HS and FF respectively, and while their technique was not widely adopted, it provided the first formulation of a content coding system in which HS and FF were independently specified.

Several of McClelland's collaborators went on to develop another projective measure of the motive to avoid failure, one that placed emphasis on external threats to achievement. Birney, Burdick, and Teevan (1969) argued that, in accordance with Atkinson's expectancy-value theory (Atkinson, 1957), individuals who set either very ambitious or relatively easy goals in level of aspiration tasks are engaging in defensive behavior in order to mitigate possible threats to self-esteem resulting from failure, and that such individuals should therefore be high in FF. Qualitative analysis of stories written by such individuals revealed a preponderance of threatening imagery: references to external blocks or hindrances (presses to use Murray's original terminology) that thwarted an individual's need for achievement. This finding led to the development of the Hostile

Press (HP) system for assessing the motive to avoid failure (Birney et al., 1969), a measure that was subsequently validated in several studies of task performance under conditions of negative social evaluation. HP was found to increase under such arousal conditions, while individuals dispositionally high in HP frequently sought to avoid achievement situations entirely.

Although the HP system has been used relatively frequently by researchers in the United States (e.g., Gelbort & Winer, 1985, Elliot & Sheldon, 1997, Fried-Buchalter, 1997), it has been criticized on conceptual grounds. Schultheiss and Brunstein (2005) argue that HP imagery, as specified by the Birney et al. (1969) system, frequently relates to more general fears regarding social evaluation and other threats to self-esteem, rather than the motive to avoid failure *per se*. Furthermore, these same researchers have pointed out that the HP system conflates active (e.g., trying harder to avoid failure) and passive (e.g., avoiding achievement situations entirely) strategies, and also suffers from motive contamination by both *n*Pow (criticism, threat) and *n*Aff (fear or rejection). These drawbacks limit the potential of the HP system to accurately characterize FF, and the measure seems to have fallen out of favor in recent years (Pang, 2010a).

Prior to the development of the HP measure, Moulton (1958) attempted to develop a new content coding system for the assessment of FF by reanalyzing data from Atkinson's (1953) study of achievement word recognition. Taking the inability to recall incomplete tasks as an indication of FF (in line with McClelland and Atkinson's conclusions), he sought to identify the distinguishing features of stories written by those high in FF, finding that the McClelland et al.'s negative scoring categories (Ga-, I-, G-, B_p, B_w) provided an accurate summary of these features when supplemented with negatively-

valenced versions of the neutral scoring categories N and thema, as well as a novel category encompassing criticism, referred to as hostile press. Extending this line of reasoning further, he developed a companion content coding system for HS that comprised McClelland et al.'s positive scoring categories, yielding a dual HS/FF system that was, to some extent, empirically derived. While the author made preliminary attempts to validate this new system (Moulton, 1958, pp. 567-571), it was not widely adopted (Brunstein & Heckhausen, 2008), and researchers in the McClelland/Atkinson tradition made no further attempts to develop PSE measures of FF once enthusiasm for the RAM/TAQ method took over. Moulton's contribution however, should not be overlooked, especially given the conceptual similarity between his system and the most influential HS/FF coding system of all – that of Heinz Heckhausen.

3.7. Measuring HS and FF: The Heckhausen (1963b) system.

Heckhausen specifically wanted to make the HS/FF distinction the centerpiece of his system for assessing achievement motivation, and as such, his measure is said to be theoretically rather than empirically derived (e.g., Pang 2010a, pp. 54-55). Unlike McClelland et al. (1949), who experimentally manipulated achievement motivation before contrasting the content of narratives written under conditions of high versus low arousal, Heckhausen (1963b) administered the PSE under a single set of conditions, characterized by neutral instructions, in which the experimenter made no reference to achievement, and picture cues that were highly unambiguous in their depiction of either success- or failure-related imagery (see section 2.1.2). He then divided up participants into two groups: those who set goals slightly higher than their previous performance level (indicative of success

motivation), and those who set either excessively easy or difficult targets (indicative of the motive to avoid failure), before looking for systematic differences in the content generated by these two groups. Those content coding categories that best distinguished between the two groups were ultimately chosen to be part of his nascent HS/FF coding system.

As mentioned in the previous section, the resulting system has much in common with both the Moulton (1956) and McClelland et al. (1953) measures. Heckhausen followed Moulton's general approach of dividing McClelland et al.'s subcategories into approach- and avoidance-related subcomponents, retaining such key constructs as need and thema (which were first invoked by Murray) but respecifying them in terms of HS and FF. Nonetheless there are significant differences. Heckhausen made an effort to streamline the system, dropping categories that appeared very infrequently (e.g., Nup) or failed to discriminate well between high HS and high FF scorers (e.g., Bw, Bp, success thema, doubtful imagery), while adding other theoretically relevant categories such as praise and criticism (similar to Moulton's hostile press) that proved useful in discriminating between success- and failure-motivated individuals.

Heckhausen's most significant contribution however, was perhaps more subtle. Rather than simply rewording McClelland et al.'s subcategories separately for HS and FF, he comprehensively revised their definitions to ensure that the new system was free from the motive contamination that plagued the McClelland et al. system. No longer were instances of competition-related dominance need scored for achievement imagery, nor were instances of help by others in an achievement task. Heckhausen recognized that these two categories were more closely related to *n*Pow and *n*Aff respectively, and thus removed them from the system, resulting in a more focused and valid measure of the

achievement motive.

3.7.1. Structure and specification.

Unlike the McClelland et al. system, the Heckhausen system does not require users to make a decision about whether or not the story is achievement-related before coding commences. The categories of AI, TI, and UI are entirely absent from the Heckhausen system, and the user is instructed to code each sentence on its own merits, irrespective of whether the entire story is deemed predominantly achievement-related or not, a revision that simplifies the scoring process considerably. As the coder reads through the story, each sentence is analyzed as a discrete unit. If the sentence is “related to achievement or embedded into an achievement-thematic context” (Schultheiss, 2001, p. 4), the coder proceeds to code for the various subcategories, each of which can only be scored once per story. Generally, the achievement-relatedness of a sentence is assumed if it describes typically achievement-related contexts or activities (e.g., employment, school) and features no alternate motive imagery, such as *nAff* or *nPow*. Positive and negative manifestations of the same scoring category, such as expectation of success and expectation of failure, can be scored in parallel, the only exception to this rule being the success thema and failure thema categories, which are mutually exclusive.

In total the Heckhausen system has 13 categories: six for HS and seven for FF. This asymmetry is due to the absence of a category corresponding to successful achievement-related outcomes, which was found to discriminate poorly between HS- and FF-motivated individuals during the system's development. Of the six HS categories, the first is need for success (NS), which is scored whenever a story character sets, or feels compelled to pursue, an achievement goal. These positive achievement intentions are

usually scored as wishes, hopes, or wants associated with attaining a standard of excellence. The second category is instrumental activity to achieve success (IS), which is scored whenever a character engages in instrumental activity designed to bring him or her closer to an achievement goal. Expectation of success (ES) is scored whenever a story character expects to succeed in an achievement-related activity, while praise (P) is scored whenever a story character praises, rewards, or distinguishes the achievements or performance of another story character. The positive affect (A+) subcategory is coded whenever a story character experiences positive affect, such as happiness, excitement, or pride, as a result of engaging in an achievement activity, moving closer to an achievement goal or succeeding in an achievement-related activity. The final HS subcategory, success thema (ST), is scored whenever achievement success becomes the dominant theme of the story. It is the only HS subcategory that is not scored for the presence of specific imagery in a sentence, and it is only scored at the end of the story once all other coding has been completed. A story is considered to possess a success thema as long as either NS or ES has been scored, and no failure categories other than A- or EF have been scored.

The FF categories almost entirely mirror the HS categories. Need to avoid failure (NF) is scored whenever a story character expresses a wish, need, or hope that failure can be avoided, while instrumental activity to avoid failure (IF) is scored for the presence of tangible activities directed towards to the avoidance of potential failure. Expectation of failure (EF) is scored whenever a story character expects their achievement-related endeavors to be unsuccessful, criticism (C) is scored whenever a story character criticizes the work or performance of another, and negative affect (A-) is scored whenever a character feels bad (e.g., experiences sadness, shame, guilt) because of their failure to

attain or move closer to attaining a certain achievement standard. The one additional coding category for FF is failure (F), which is coded whenever characters fail, either completely or partially, in their current achievement activities, or when the author mentions instances of past failure. A story is considered to possess a central failure theme (FT) if either NF or F have been scored, and no success categories other than IS have been scored.

Typically, scores are summed among the HS and FF categories to provide separate indices of hope of success and fear of failure. As with the McClelland et al. and Winter systems, these scores are usually summed or averaged across the picture cue battery to compute dispositional scores (Pang, 2010b). However, a unique property of the Heckhausen system is that HS and FF scores can be combined to yield two additional variables. The difference between the two scores ($HS - FF$) is referred to as net hope (NH), while the sum of the two scores is termed aggregate motivation (AM). These combined variables were used extensively by Heckhausen and colleagues in their early work, but are used only infrequently by contemporary researchers, who generally prefer to work with the pure HS and FF constructs.

3.7.2. Validation.

While the Heckhausen system was not used by researchers in the McClelland/Atkinson tradition, it was validated extensively by a core group of researchers in Germany. Interestingly, researchers (e.g., Fisch, 1970) have found little or no correlation between Heckhausen's FF and TAQ scores, further suggesting that the two measure fundamentally different constructs. In a similar vein, Halisch and Heckhausen (1988) found that neither HS nor FF scores correlated with questionnaire measures of

achievement motivation, supporting the implicit/explicit motive distinction and providing evidence of divergent validity in the Heckhausen system.

Generally, HS motivated individuals have been found to behave in a manner very similar to those motivated by *n*Ach, which is unsurprising given the emphasis placed on success in the original McClelland et al. system. Their research has also managed to paint a clearer picture of those motivated by FF. Schneider (1978) found that individuals scoring highly for Heckhausen's FF avoided tasks of moderate difficulty, a finding that is consistent both with Atkinson's expectancy-value theory and with findings in the RAM/TAQ tradition. High FF scorers also tend to recall tasks that they have completed better than those that they have failed to complete (Heckhausen, 1980), a finding that resonates with McClelland, Atkinson and colleagues' earlier attempts to characterize FF in the tripartite division paradigm. Further research conducted by Heckhausen and colleagues elaborated on this picture, characterizing those motivated by FF as worry prone and cautious, qualities that could occasionally (but not always) impair performance. Consistent with this account, Heckhausen (1980) reported that individuals high in FF took longer to do their homework and performed worse under time pressure compared to those motivated by HS. Critically, Heckhausen (1980) found that FF-motivated participants performed better after success feedback, an early finding that highlights the importance of identifying the conditions under which fearful individuals can thrive. Further evidence supporting the validity of the Heckhausen system comes from studies investigating other psychological consequences of the HS/FF distinction. In a study of perceptions of the structure and experience of time, HS-motivated individuals, as well as those characterized by high AM, were found to structure time spans further into the future in than those

motivated by FF (Heckhausen, 1963a). Others have found differences in memory. Pang et al. (2009) found that HS-motivated individuals have better memory for successful peers, while those motivated by FF recall unsuccessful peers more accurately.

3.7.3. Weaknesses.

Heckhausen's coding system addressed a major weakness of the McClelland et al. system by separating the measurement of the hope and failure motives, acknowledging the fundamental distinction between the approach- and avoidance-related manifestations of the achievement motive. However the Heckhausen system falls short of McClelland et al.'s system in the manner of its derivation. While McClelland and colleagues derived their system empirically by arousing the motive to varying degrees and contrasting the content of generated narratives, Heckhausen's system was formulated with the explicit goal of characterizing and maintaining a functional difference between HS and FF. The validity and necessity of the HS/FF distinction was an *a priori* assumption in the Heckhausen system, meaning that the development of the system was driven primarily by theory in a top-down fashion. Pang (2010a) has argued that empirically derived coding systems should exhibit better construct validity than systems shaped by literature review and deductive reasoning, since the former systems are based on the actual effects of motivation on fantasy (Pang, 2010a, p. 55). In seeking only to respecify McClelland et al.'s extant subcategories in terms of HS and FF, Heckhausen may have missed out on important subcategories that distinguished between high HS and high FF scorers. Such omissions may limit the validity of the system if important facets of the target motive are not assessed by the measurement tool.

3.8. Measuring HS and FF: The Pang (2006) system.

Recently, Pang (2006) has developed another content coding system for the assessment of hope of success and fear of failure. In developing her revised system, she attempted to address the major shortcoming of Heckhausen's HS/FF coding system: the lack of empirical derivation. To this end, she conducted a study in which participants were exposed to either positive self-referenced feedback, negative self-referenced feedback, or no feedback regarding their performance in a challenging mental concentration task. Self-referenced feedback has been found to be potent in arousing the achievement motive (Brunstein & Maier, 2005; Brunstein & Schmitt, 2004), and so Pang reasoned that exposure to intermittent success-related feedback ("you performed significantly better than on your previous trials!") should arouse HS in participants, whereas exposure to intermittent failure-related feedback ("you performed significantly worse than on your previous trials!") should arouse FF. Participants in the no feedback condition were given no feedback on their performance at all. Consistent with the Heckhausen conceptualization, it was assumed that the arousal of both HS and FF would be activating with respect to achievement behavior, with participants working harder to obtain incentivizing feedback in the former condition and to avoid disincentivizing feedback in the latter. This methodology represented a replication and extension of McClelland et al.'s (1949) procedure, as HS and FF were experimentally aroused and their projective consequences compared, both with each other and with a state of low arousal. Further replicating McClelland et al.'s methods, she then used qualitative analytic procedures derived from grounded theory (Glaser & Strauss, 1967) to identify the content coding categories that differentiated narratives written under the three conditions of motive

arousal. These differences were subsequently reified in Pang's (2006) *Revised Content Coding Manual for Hope of Success (HS) and Fear of Failure (FF)*.

3.8.1. Structure and specification.

The Pang system has much in common with the Heckhausen system despite the fundamental difference in their methods of derivation. Like the Heckhausen system, the Pang system is divided into HS and FF subscales and features a total of 13 coding categories, although in the Pang system the asymmetry is reversed, with HS encompassing seven subcategories and FF encompassing six. In identifying the most salient discriminators, Pang dropped several subcategories from the Heckhausen system that were ineffective in discriminating between the high and low arousal conditions, such as thema (ST, FT), need to avoid failure (NF), instrumental activity (IS, IF), negative affect (A-) and criticism (C), while also including a number of novel categories that were more effective, some of which resemble those used by McClelland et al. (1953) while others are entirely new. The most important innovation of the Pang system however, was inspired by the Winter system. Unlike the Heckhausen system, where each subcategory can only be scored once per story, the Pang system allows for each subcategory to be scored multiple times in a single story. The only restriction is that each category can only be scored once per sentence (as in the Heckhausen system, the sentence is the unit of analysis).

The first category in Pang's HS subscale is positive achievement goal (PAG), which is conceptually similar to Heckhausen's NS and McClelland et al.'s N. PAG is scored whenever a story character "expresses or experiences an intention (desire, wish, need) to fulfill a positive achievement goal" (p. 5). Unlike previous systems, the Pang

system has two positive affect categories: energizing emotion (EE), which is scored whenever a positive emotion promotes further engagement in achievement-related activity, and positive emotion (E+) which is scored whenever a story character experiences positive affect as a result of goal attainment or progress. Self-regulation (SR) is scored for mention of preparation and training, self-scrutiny, or hard work, while persistence (PER) codes for instances of compensatory effort (after encountering a setback) or strain/self-sacrifice in the pursuit of an achievement goal. Expectation of success (ES) and praise (P) are retained from the Heckhausen system, while success outcome (S) is scored whenever a positive achievement goal is fulfilled.

Unlike in the Heckhausen system, there is no avoidant achievement need category in the Pang system. As such, the first FF category is inhibiting emotion (IE), which is scored whenever a story character experiences a negative emotional state that inhibits or prevents their engagement in achievement-related activity. The second FF category, quality control by others (QCO), encompasses instances of monitoring of a story character's work by another story character. Obstacle (O) is similar to McClelland et al.'s Bw, and is scored whenever external factors disrupt a character's progress towards an achievement goal, while expectation of failure (EF) and failure (F) are largely unaltered from the Heckhausen system. The final failure category, life dream (LD), is scored whenever an achievement goal is described as a life-long goal or ambition that is central to the character's self-concept.

3.8.2. Weaknesses.

The weakness of the Pang (2006) system lies largely in a lack of validation. While her system addresses the major limitations of the aforementioned alternatives (e.g.,

separation of HS and FF, empirical derivation), it has yet to be adopted by the wider research community, and so few studies have demonstrated significant prediction of relevant criterion variables. Pang (2010a) has provided preliminary support, demonstrating that HS and FF scores predict differential responding in both a risk-taking measure, the Iowa Gambling Task (IGT; Bechara, Damasio, Damasio, & Anderson, 1994), and in a commonly-used concentration task, the d2 test of attention (Brickenkamp & Zillmer, 1998). She also found that her HS and FF scores do not correlate with explicit measures of these constructs, providing evidence of divergent validity..

3.9. The need for a consolidated coding system.

The preceding sections have detailed four prominent systems for the PSE assessment of implicit achievement motivation. The existence of so many alternatives presents researchers with a difficult choice. Each system differs in terms of its derivation (theoretical vs. empirical vs. a combination of the two), its intended motivational target(s) (*n*Ach vs. HS/FF), and in the specific items (i.e., coding categories) that contribute to, and hence define, the motives of interest. Each has its own strengths and weaknesses, and to date no comprehensively validated system has addressed all the shortcomings of the most widely used coding manuals. Furthermore, it is well-documented that PSE research is both time consuming and labor-intensive (e.g., Vane, 1981), and researchers rarely have the inclination to master multiple coding methodologies. As such, the implicit achievement motive literature is fragmented through use of different instruments, quite possibly measuring different motivational constructs, the exact relationships between which have not been comprehensively delineated. Findings from research conducted using

one coding instrument may be difficult to integrate into the body of evidence derived from use of another, since the degree of correspondence between two constructs (e.g., Heckausen's HS and Winter's *n*Ach) has not been thoroughly investigated.

These facts speak of an urgent need to reconcile the separate research traditions through development of an all-encompassing composite instrument that retains the most informative elements of each constituent coding system while eliminating redundancy wherever possible. Not only will this improvement facilitate exchange between researchers in different implicit motive assessment traditions, but it also has the potential to increase integration with the psychological mainstream, something that some authors have noted to be lacking (e.g., Woike, 2008). The PSE and the associated field of implicit motive research are misunderstood by many contemporary personality psychologists, who associate them with the unscientific approach of the psychoanalytic tradition (Woike & McAdams, 2001). As such, there have been few attempts to utilize PSE assessment by researchers outside the core implicit motive research tradition, who may forego valuable insight by failing to acknowledge the implicit-explicit distinction. The composite system must build on the work of Pang (2006), addressing the weaknesses associated with previous systems while giving rise to valid *n*Ach, HS, and FF scores that predict relevant criterion variables. The development of such an instrument should encourage more researchers to adopt PSE methodology, enhancing our understanding of HS, FF, and *n*Ach.

3.10. The present research.

The second half of this thesis describes the development of the first consolidated

coding system for the simultaneous assessment of *n*Ach, HS, and FF. To this end, PSE narratives were content coded using the four predominant systems (Heckhausen, 1963b; McClelland et al., 1953; Pang, 2006; and Winter, 1994), and cluster analytic procedures were used to assess relatedness among both individual categories and the higher-order motives of *n*Ach, HS, and FF. Clustering solutions were then used to consolidate the four systems into a single composite system, which was subsequently validated through prediction of several theoretically relevant behavioral variables. The contributions of the new system are:

- a) Incorporation of HS/FF Distinction: The new system addresses the most significant weakness of the commonly used McClelland et al. and Winter systems by separately assessing the approach- and avoidance-related subcomponents of the achievement motive, while also measuring the *n*Ach to facilitate integration with existing achievement motivation literature and encourage adoption by researchers working in the McClelland/Atkinson and Winter traditions.
- b) Characterization of *n*Ach, HS, and FF: By virtue of its derivation, the new system confirms the equivalence of *n*Ach and HS, and confirms the distinctiveness of the FF from these approach-related motives.
- c) Simplification of Administration: The new system is as parsimonious as possible, using the smallest number of possible categories to provide valid measures of *n*Ach, HS, and FF. Categories retained for use in the composite system were redefined using the simplest possible language, and formalized into a coding manual that replicates the simplicity of the Winter system
- d) Avoidance of Motive Contamination: The consolidated system also avoids the

motive contamination that plagued the original McClelland et al. system by discarding categories that were found to associate with alternative motives assessed using the Winter system. Categories that were found to cluster with either the *nAff* or *nPow* constructs during the analysis were deemed unsuitable for use in the finalized *nAch* coding system and consequently discarded.

STUDY 1

4. Study 1 Overview

The first study attempted to construct and validate an entirely new picture set for the assessment of the approach- and avoidance-related manifestations of the achievement motive; a picture battery that overcomes the limitations of current picture sets identified in the introduction. Specifically, the study was designed to produce a picture set that offers improvement in the five key domains outlined in section 2.6:

- a) Improved relevance and decreased bias
- b) Suitability for the assessment of HS and FF
- c) Independent validation
- d) Systematic comparison
- e) Appropriate set ambiguity

Study 1 comprised two distinct phases. In study 1A, a novel achievement picture set was constructed through pre-testing and selection of a large number of candidate images, with only the most suitable and effective images being retained for further testing. Study 1B sought to establish the psychometric fitness of this new picture set, assessing

test-retest reliability, divergent validity, and predictive validity, while comparing the performance of the new picture set to that of a comparison picture set comprising currently used images.

The methods adopted in study 1A were designed to offer significant improvement in domains (a) and (b). Firstly, images were selected from a highly diverse and gender-balanced pool of candidate images featuring a variety of achievement-related activities and settings. Consequently, it was intended that this resulting picture set would exhibit greater diversity of gender, ethnicity, and achievement setting than two commonly used picture sets: McClelland and colleagues' (1953) standard set and Heckhausen's (1963b) HS and FF picture set. It was also expected that this enhancement of the universality and relevance of the new picture set would afford significant improvements in validity, which was to be assessed in study 1B. Secondly, candidate images were retained for inclusion in the new picture set on the basis of their ability to pull simultaneously for both HS and FF, which ensured that the selected images exhibited sufficient cue strength and cue ambiguity for the valid assessment of both HS and FF. Adoption of this criterion stands in contrast to many widely used picture sets, whose suitability for the assessment of HS and FF has never been critically examined (e.g., the McClelland et al. standard set). It was further anticipated that both of these improvements would manifest in enhanced validity in study 1B.

Study 1B was designed so as to offer evidence of improvement in domains (c) and (d). Motive scores derived from the new picture set were systematically compared to those derived from currently-used images using an independent coding system:

Heckhausen's (1963b) need achievement measure. It was hypothesized that scores derived from the new picture set would display good test-retest reliability, convergent validity, and divergent validity, and that they would display better predictive validity than those derived from the comparison picture set.

In order to assess point (e)—the importance of set ambiguity—two novel picture sets were constructed and validated in parallel. Two variants, one that was relatively ambiguous and one that was unambiguous in its depiction of achievement-related activity, were firstly constructed during study 1A. The unambiguous picture set included the eight most effective pictures (with respect to pulling for both HS and FF) identified during pretesting, while the ambiguous picture set included the six most effective pretested images and two additional images that pulled preferentially for alternate motives: *nAff* and *nPow*. Of these two novel picture sets, it was expected that the ambiguous novel picture set would offer the greatest improvements in validity over the commonly used pictures, a hypothesis that was tested in study 1B.

4.1. Hypotheses.

Hence, the following hypotheses were made:

4.1.1. Test-retest reliability.

It was hypothesized that scores derived from the ambiguous novel picture set would display a one-week stability coefficient of about .60 (H1.1), in line with Schultheiss and Pang's (2007) meta-analysis of studies assessing PSE test-retest reliability. Such an instrument would be said to possess good test-retest reliability. Furthermore, it was hypothesized that scores derived from the ambiguous novel picture

set would display greater test-retest reliability than those derived from the unambiguous novel picture set (H1.2). The one-week stability coefficient was calculated by correlating the scores obtained in an initial administration with those obtained during an identical administration one week later.

4.1.2. Divergent validity.

Implicit measures of motivation have been shown to associate weakly with more declarative, explicit motive assessments (McClelland, Koestner, & Weinberger, 1989), and so a valid measure of implicit achievement motivation should not correlate with its explicit counterparts. As such, it was hypothesized that the scores derived from the ambiguous picture set would not correlate with scores derived from two commonly used explicit measures of the achievement motive: the achievement subscale of Jackson's Personality Research Form (PRF; 1974) and the Hope of Success and Fear of Failure scale (Schultheiss & Murray, 2002) (H2). It was also anticipated that the scores derived from the unambiguous picture set would not correlate with either PRF-derived *sanAch* or with their explicit analogue from the Hope of Success and Fear of Failure scale, since there is no grounds for suspecting that the unambiguous picture set constitutes a true explicit measure. As such, the divergent validity hypothesis pertains to the ambiguous picture set only.

4.1.3. Predictive validity.

Finally, it was also hypothesized that scores derived from the ambiguous novel picture set would better predict theoretically relevant behaviors in two tasks—the Wisconsin Card Sorting Test (WCST; Grant & Berg, 1948) and the Balloon Analogue

Risk Task (BART; Lejuez et al., 2002)—than those derived from either the unambiguous novel picture set or the set of commonly used picture cues.

The rationale for the selection of these two tasks is as follows. Given that good performance in the WCST is contingent on both speed and accuracy, it was reasoned that the approach orientation of high HS scorers should manifest as a desire to maximize positive outcomes rather than minimize negative outcomes, leading to greater speed at the expense of accuracy. Such behavior would be consistent with Schultheiss and Brunstein's (2005) suggestion that HS-motivated individuals exhibit "greater tolerance for frustrations" (p. 11), in that they are less concerned about the possibility of encountering momentary frustrations and temporary setbacks in their pursuit of good performance. In the WCST, this quality should manifest as a willingness to make a few mistakes as long as overall goal pursuit (a generally high level of speed and accuracy) is not compromised. As such, it was hypothesized that HS scores from the novel ambiguous picture set would positively predict sorting speed (H3.1) and negatively predict sorting accuracy (H3.2), whereas HS scores derived from either the unambiguous picture or the set of commonly-used pictures would not.

In the case of the BART, a measure of risk-taking behavior, it was hypothesized that FF scores derived from the novel ambiguous picture set would negatively predict both risk-taking (H4.1) and task speed (H4.2), whereas FF scores derived from either the unambiguous novel picture set or the set of commonly-used images would not. Since the possibility of failure in the BART can only be mitigated by engaging in less risky behavior, it was predicted that scores from the most valid picture set would negatively

predict risk taking. Additionally, since performance in the BART is time-independent, it was further reasoned that scores from a valid FF measure should negatively predict task speed, since FF scorers are afforded the chance to approach the task in a more cautious and deliberative manner. Heckhausen (1980) has documented the tendency of FF-motivated individuals to spend longer on homework assignments to ensure good performance. Since performance in the BART is not judged in terms of speed, it was reasoned that this task should allow FF-motivated individuals to indulge this natural tendency for caution and deliberation. This quality of the BART contrasts with the definition of good performance in the WCST, in which rapid responding is explicitly requested, thus denying FF-motivated individuals the opportunity to approach the task in the cautious manner that suits them best. It was therefore predicted that FF scores from the ambiguous novel picture set would predict slower BART speed whereas FF scores from either the unambiguous novel picture set or the comparison picture set would not.

5. Study 1A: Developing the Picture Set

5.1. Method.

5.1.1. Participants.

All participants ($N = 186$) were undergraduate students from a large Singaporean university, who completed the assigned tasks in exchange for partial course credit. All study elements were administered in English, which is the university's teaching language as well as the vernacular in Singapore. The sample comprised 124 females and 62 males, and had a mean age of 21.1 years ($SD = 1.6$). 87.1% of the participants were ethnically

Chinese, 4.3% were Malay, 5.4% were ethnically Indian, while the remaining 3.2% described themselves as “other”. All participants provided informed consent, and were fully debriefed at the end of the experimental session.

5.1.2. Design and procedure.

5.1.2.1. Pretest round 1.

Study 1A involved four separate rounds of pretesting and selection. The first round involved the construction of an initial cohort of 50 royalty-free candidate images purchased from an online vendor of stock photography. I chose these images with a view to minimizing the race and gender bias that exists in other picture sets, as well as to improve extensivity with respect to different achievement settings. Candidate images were chosen on the basis of perceived verisimilitude and the presence of one or more actors who appeared to be engaged in goal-directed behavior within an achievement-related context. Achievement-related contexts were identified from a comprehensive survey of the extant literature on PSE-measured achievement motives. It was established that workplace, academic, and competitive sporting environments were the most widely studied achievement contexts. Additionally, images from a novel category of achievement-related contexts, referred to as *creative endeavor*, were also included so as to compensate for the under-representation of creative pursuits (e.g., culinary arts, visual arts) in the achievement motive literature. Images categorized as creative endeavor depicted individuals involved in more creative achievement-related activities, with examples of such images including a chef meticulously carving fruit and an artist putting the finishing touches to an impressionist painting.

5.1.2.2. Pretest round 2

A second round of pretesting was conducted in order to further improve the diversity of demographic characteristics represented in the candidate image cohort, and to eliminate images judged to possess limited face validity for PSE assessment of achievement motivation. Initially, the size of the candidate image pool remained constant at 50 images. Firstly, content frequencies and percentages were calculated for the first cohort in order to quantify the relative prevalence of different genders, ethnic groups, and achievement settings. For example, the number of male and female characters depicted was summed across the entire image cohort so that the percentage of male and female characters in the whole image cohort could be calculated. Deviation from a 50:50 split was taken as an indication of gender imbalance within the entire set of candidate images. On the basis of these frequencies and percentages, images were successively replaced until an improved cohort with more balanced demographic percentages had been assembled (see table 1 for details of demographic frequencies in the first and second cohorts). Subsequently, this improved cohort was rated for achievement-relatedness (i.e., face validity) by an independent expert trained in the administration of several PSE measures of *n*Ach, HS, and FF. Those that were deemed to exhibit low face validity were dropped, leaving a reduced cohort of 28 images that were carried forward to the third round of pretesting.

Table 1

Demographic and achievement setting comparisons for the pretest cohorts, new picture set variants, and comparison picture sets																		
	Pretest Cohort 1			Pretest Cohort 2			Unambiguous Set			Ambiguous Set			McClelland et al.			Heckhausen		
	Freq.	%		Freq.	%		Freq.	%		Freq.	%		Freq.	%		Freq.	%	
<i>Gender</i>																		
Tot. Male:	54	63.5%		47	54.7%		10	71.4%		12	70.6%		15	100.0%		9	100.0%	
Tot. Female:	31	36.5%		39	45.3%		4	28.6%		5	29.4%		0	0.0%		0	0.0%	
Pri. Male:	11	44.0%		11	37.9%		8	66.7%		6	54.5%		11	100.0%		9	100.0%	
Pri. Female:	14	56.0%		18	62.1%		4	33.3%		5	45.5%		0	0.0%		0	0.0%	
<i>Race</i>																		
Caucasian:	60	70.5%		56	65.0%		5	35.7%		10	58.8%		15	100.0%		9	100.0%	
East Asian:	15	17.6%		17	19.8%		2	14.3%		1	5.8%		0	0.0%		0	0.0%	
South Asian:	2	2.3%		3	3.5%		1	7.1%		0	0.0%		0	0.0%		0	0.0%	
Mid. Eastern:	0	0.0%		0	0.0%		0	0.0%		0	0.0%		0	0.0%		0	0.0%	
Hispanic:	1	1.1%		0	0.0%		0	0.0%		0	0.0%		0	0.0%		0	0.0%	
Black:	1	1.1%		7	8.1%		4	28.6%		3	17.7%		0	0.0%		0	0.0%	
Ambiguous:	6	7.1%		3	3.5%		2	14.3%		3	17.7%		0	0.0%		0	0.0%	
<i>Setting</i>																		
Work:	30	60.0%		22	44.0%		2	25.0%		2	25.0%		5	63.5%		4	66.7%	
Academic:	8	16.0%		10	20.0%		2	25.0%		2	25.0%		1	12.5%		2	33.3%	
Sport:	7	14.0%		13	26.0%		4	50.0%		2	25.0%		0	0.0%		0	0.0%	
Creative:	4	8.0%		5	10.0%		0	0.0%		0	0.0%		1	12.5%		0	0.0%	
Other:	1	2.0%		0	0.0%		0	0.0%		2	25.0%		1	12.5%		0	0.0%	

Note. Primary characters are those who are most prominent in the image. Images may have more than one primary character if two characters are emphasized equally. Settings designated as “other” are either unrelated to achievement or ambiguous in their depiction of motive-relevant activity.

5.1.2.3. Pretest round 3

In the third round of pretesting, 100 participants judged the achievement-relatedness of the 28-image cohort derived during round two using a simplified approximation of PSE methodology. Participants were asked to view each of the images before writing down strings of adjectives and other descriptive words or phrases relating to their interpretation of the image, as opposed to the narratives that would be requested in a conventional PSE administration. This method greatly reduced the time taken for pretesting since participants were able to respond to all 28 images over the course of a single experimental session, whereas a conventional PSE is usually limited to an eight image maximum due to concerns over fatigue and decreasing validity (Reitman & Atkinson, 1958). Before responding to each image participants were provided with the following instructions: "Use the box below to write down single words describing the picture below. Please write down only single words, and do your best to list as many as you can in the time allotted. Words should be related to the people in the picture - what they are doing, thinking, feeling and wishing for." The final sentence closely matches instructions typically provided during conventional PSE administration, and was adapted from Schultheiss and Pang (2007, p. 333). Each participant was required to generate a minimum of 10 words per image, up to a maximum number of 20. No limit was imposed on the amount of time participants could spend responding to individual images, although participants were instructed to spend no more than one hour responding to the entire cohort of 28 images. On average, each participant generated 10.56 words in response to each image. Examples of responses to the image *dropped papers*, which depicted an

elderly man stooping to retrieve a pile of documents from the floor, included the words “messy”, “stressed”, “puzzled”, and “perfection”.

The resultant adjective words and phrases were then coded for achievement-related imagery by two independent raters using a binary coding system in which each word/phrase could be classified either as achievement-related or not achievement-related. Both raters were trained in the McClelland et al. (1953), Winter (1994), Heckhausen (1963b), and Pang (2006) systems, and classifications were made in a conservative fashion such that any word that was deemed to be either unrelated to achievement or ambiguous with regard to its motive associations was designated not achievement-related. Words such as “ace”, “careful”, and “masterpiece” were classified as achievement-related, whereas words such as “arrogant”, “complications”, and “moody” were not. Once 100% inter-rater agreement on the achievement-related words had been established, an achievement-relatedness score was calculated for each image on the basis of how many times achievement imagery had been elicited across the entire pilot test. This technique allowed for image ranking on the basis of achievement-relatedness, with all those images eliciting 25% or more achievement-related words being selected for the final round of pretesting. Applying this rationale, 18 images were carried forward to the final pretest round while the remaining 10 were discarded. The mean percentage of achievement-related words for the retained group was 34.30 ($SD = 6.29$), while the mean for the discarded group was 18.91 ($SD = 4.43$). This difference was statistically significant, $t(26) = 6.83, p < .001$.

5.1.2.4. Pretest round 4.

In the fourth and final round of pretesting, the 18 retained images were assessed using full PSE methodology. Eighty-six participants wrote PSE stories in response to an eight-picture battery that included a subset of the candidate images, and the resulting protocols were scored for HS and FF by two raters trained in the application of the Heckhausen (1963b) coding system. The final eight images were selected on the basis of motive pull and cue ambiguity. Firstly, any image that did not demonstrate significant motive pull (more than 50% of candidates to write narratives containing at least one instance of achievement-related imagery; Schultheiss & Brunstein, 2001) for either HS or FF was discarded. Secondly, the mean number of instances of both HS- and FF-related imagery per thousand words (across all participants) was calculated for each of the remaining images. These indices of HS and FF pull were then consolidated into a third variable measuring cue ambiguity by taking the larger of the two numbers and dividing by the smaller. As a result, highly ambiguous images had a cue ambiguity value approaching one while unambiguous images with much higher pull in one category than the other gave rise to a larger number. Images were then ranked smallest to largest in terms of their cue ambiguity value, with the eight smallest (and most ambiguous) being retained.

The eight images with the highest cue ambiguity for HS/FF (*basketball, blueprints, forehead, lecture theater, operating theater, skaters, squash, and student*) formed the unambiguous (with respect to *nAch* depiction vs. other motives) variant of the novel picture set (i.e., the set with low set ambiguity). The ambiguous variant with high set ambiguity comprised the six pictures with the highest cue ambiguity for HS/FF

(*basketball, blueprints, lecture theater, operating theater, skaters, and student*) and two images that have previously been shown to pull for an alternate motive (Pang, 2010b): *couple by river* (*nAff*) and *hooligan attack* (*nPow*), which replaced the two least ambiguous (i.e., least effective) HS/FF images *forehead* and *squash*. As such, the ambiguous and unambiguous picture sets differed only in terms of two images. Reproductions of the images comprising the ambiguous and unambiguous picture sets can be found in Appendices A and B respectively.

5.2. Results.

5.2.1. Diversity.

Both the ambiguous and unambiguous variants of the new picture set displayed diversity in terms of gender, ethnicity, and achievement setting (see table 1 for details). While the two picture sets still feature more men than women, both represent a substantial improvement over historic, frequently used picture sets such as the McClelland et al. (1953) standard set or the Heckhausen (1963b) picture set, neither of which depict any women at all. Furthermore, the gender balance of the primary characters (i.e., individuals who are the main subject of the image, rather than just background characters) in both the ambiguous and unambiguous variants approaches equality, particularly in the ambiguous variant. The ambiguous and unambiguous picture sets also feature ethnic diversity, with four and three ethnic categories being represented respectively. This composition compares favorably with both the McClelland et al. (1953) standard set and the Heckhausen (1963b) set, which exclusively depict Caucasian characters. Finally, the extensities of both the ambiguous and unambiguous picture set variants are superior to

those of the commonly used sets, with work, sporting, and academic achievement settings all featured. The comparison sets predominantly feature work settings, and do not feature any sporting scenarios.

5.2.2. Cue strength and cue ambiguity.

Each of the images retained during the fourth and final round of pre-testing elicited at least one instance of motive-relevant imagery in at least 50% of participants (see figure 2), and so can be said to pull significantly for *nAch*. Five of the eight images pulled significantly for both HS and FF, while two (operating theater and squash) only exhibited significant pull for HS. *Forehead* pulled significantly for FF only. The retained images also exhibited good cue ambiguity. The most ambiguous image, *blueprints*, had a cue ambiguity value of 1.01 (indicating that it pulls equally for HS and FF, whereas the least ambiguous of the retained images, *forehead*, had a cue ambiguity value of 2.35, indicating that it was more than twice as effective in pulling for FF than HS. The remaining images had cue ambiguity values that fell between these two extremes.

5.3 Discussion.

The rationale and selection criteria applied in the first and second rounds of pretesting succeeded in creating a pool of candidate images that were highly diverse in their depiction of race, gender, and achievement setting. For example, the second pre-test cohort featured an almost equal number of male and female actors, and several images in the creative endeavor category. However, after the additional criteria of cue ambiguity and adequate motive pull were imposed during the third and fourth rounds of pretesting, the

diversity of the final ambiguous and unambiguous picture set variants was somewhat reduced. Nonetheless, the final picture set variants still represent a marked improvement over commonly used picture sets such as the McClelland et al. (1953) standard set and the Heckhausen (1963b) picture set, both of which depict only Caucasian males and focus predominantly on employment settings. As such, the aim of constructing a picture set with enhanced diversity was fulfilled, addressing the concerns raised in section 2.4.2 of the introduction. While it is perhaps unfortunate that some of the diversity achieved in the earlier cohorts was lost during the later rounds of selection, the quest for diversity (which was expected to confer improvements in relevance, universality, extensity, and ultimately validity) must be balanced with a requirement for selecting images with the correct balance of cue strength and cue ambiguity, the satisfaction of which is discussed presently.

Figure 2

Achievement cue strength of the pretested images

Note. Bask = basketball, Blue = blueprints, Fore = forehead, Lect = lecture theater, Oper = operating theater, Skate = skaters, Sqsh = squash, Stdnt = student

The image selection criteria applied in the third and fourth rounds of pretesting were designed to give rise to a picture set comprising images that exhibited an appropriate balance of cue strength and cue ambiguity with respect to HS and FF, addressing the concerns raised in section 2.4.3. The pretesting procedures were successful on both counts. Each of the images were found to pull significantly for at least one of the two motives, with five of the eight pulling significantly for both HS and FF. In addition, the selection of only the most ambiguous images during the final pretesting round ensured that none of the images pull overwhelmingly for either HS or FF at the expense of the other. While the less ambiguous of the selected images, such as *forehead*, still pulled twice as strongly for one motive than the other, this still represents an improvement over the only other picture set specifically designed for HS and FF assessment—the Heckhausen (1963b) set—since Heckhausen's images were specifically chosen to evoke either HS or FF exclusively. As such, the aim of assembling a picture set containing images with the appropriate balance of cue strength and cue ambiguity was fulfilled.

Ultimately, the construction of picture sets exhibiting enhanced diversity in combination with satisfactory pull and cue ambiguity with respect to HS and FF was not an end in and of itself. These improvements in the properties of individual picture cues (pull, cue ambiguity, universality, relevance) and the overall picture sets (diversity, extensity) were intended to give rise to a more valid PSE assessment instrument. Having established the suitability of the novel images for the assessment of HS and FF, study 1B then sought to investigate whether these improvements in diversity, in conjunction with

satisfactory stimulus pull and cue ambiguity, would manifest in enhanced validity, particularly in the case of the ambiguous variant.

6. Study 1B: Validating the Picture Set

6.1. Method.

6.1.1. Participants.

All participants ($N = 80$) were undergraduates from the same university as study 1A. The sample was 54.7% female and 80.0% ethnically Chinese with a mean age of 21.52 years ($SD = 1.83$). 4.0% participants were Malay, 10.7% were ethnically Indian, while the remaining 5.3% described themselves as belonging to an unlisted ethnic group. All participants were recruited from the university's introductory psychology class. Data for six participants were excluded, either in light of their failure to complete one or more of the experimental sessions or due to their generation of extremely brief PSE stories that precluded valid scoring for motive content. This resulted in a sample of 74 participants (54.2% female) who contributed data to the final analysis. Participants were rewarded with partial course credit, and informed consent was obtained in all cases.

6.1.2. Design and procedure.

Participants were randomly assigned to one of two experimental groups: ambiguous ($n = 41$) or unambiguous ($n = 33$). Participants in both groups completed two kinds of PSE: one using either the ambiguous or unambiguous variant of the new picture set, and one comprising a set of widely used images. Comparison of the scores derived from these two kinds of picture set directly tested the hypothesis that scores derived from the novel picture sets, particularly the ambiguous picture set, offered improvements in predictive validity over the widely used pictures.

Participants in both conditions completed a series of four experimental sessions, each one lasting around half an hour, over a period of ten days. Session one consisted of a PSE administration of either the ambiguous or unambiguous picture set (depending on group allocation), while session two was a straightforward replication of session one that took place exactly one week later. Session three, which featured the validation tasks, took place on the day following session two. Finally, the fourth session, which involved the administration of the comparison image PSE as well as demographic items, was administered one day after session three. Participants were fully debriefed after the final experimental session.

The entire study was administered individually and remotely, using the web version of the popular experimentation software Inquisit 3.0 (2010) in combination with the online survey utility Qualtrics (2011-2013). Participants were asked to complete the online PSE either from home or using the university's computer facilities. A secure link to the online experiment was emailed to all participants around half an hour prior to the pre-agreed commencement time, along with detailed instructions regarding test conditions and task preparation, as well as a brief trouble-shooting guide. Participants were instructed to complete the task alone and to refrain from engaging in any other activities (e.g., internet browsing, listening to music) while the experiment was in progress. An experimenter was available during each administration to answer queries either by phone or email. Once the participants had completed a session, their data were automatically uploaded to a secure online database for later retrieval by the experimenter.

Prior to commencing each PSE, participants received the same standardized instructions adapted from Schultheiss and Pang (2007). These instructions have been successfully used in a number of studies and have been compiled from various sources (Atkinson, 1958; Lundy, 1988; Smith, 1992). Also in accordance with Schultheiss and Pang (2007), guiding questions, which referenced these instructions, remained visible in the top left hand corner of the screen throughout the experimental session. After acknowledging that they had read and understood the instructions, participants were shown each image for 10 seconds, after which the image disappeared and they were given four minutes to write an imaginative story. Once the four-minute time limit had been reached, participants were prompted to proceed to the next image. Image presentation order was randomized for all participants.

All PSE protocols were coded for HS and FF by two independent raters using the English translation of Heckhausen's (1963b) measure (Schultheiss, 2001). Use of the Heckhausen system allowed for validation of the picture independent of an associated coding system, a weakness of commonly used picture sets that was identified in the introduction. The raters, both of whom had previously reached a level of 85% agreement with the Heckhausen manual's training materials, first coded 10% of the dataset and established an inter-rater reliability of >85% before proceeding to code the rest of the data. The final concordance rate for the two scorers across the entire dataset was 99.7% after all initial coding disagreements were resolved by rater discussion. The few remaining disagreements were resolved by averaging counts across the two differing scores. Average story length and frequency of motive imagery were significantly

correlated ($r = .13, p = .003$), so the final scores for each story were then corrected for word count using a procedure recommended by Pang (2010b), whereby the total picture-specific motive scores for HS and FF are multiplied by 1000 and then divided by word count, which restates the scores as the number of instances of motive-relevant imagery per 1000 words. These word count corrected HS and FF scores were then averaged across the eight stories in each experimental session to provide mean HS and FF scores, which were entered into the subsequent analyses as indicators of hope of success and fear of failure.

6.1.3. Materials.

6.1.3.1. Novel picture cues.

As described in the previous section, the unambiguous novel picture set comprised the eight pre-tested images found to be most effective in pulling for both HS and FF, while the ambiguous picture set comprised the six most effective of these images and two additional images previously shown to pull preferentially for alternate motives. More details of these two picture sets can be found in section 5.1.2.4.

6.1.3.2. Comparison picture cues.

The pictures used in the comparison PSE were taken from a variety of sources, although all have been widely used in achievement motive research. The rationale behind image selection here was to adequately represent each of the picture selection traditions outlined in section 2.3, meaning that pictures from the McClelland and Heckhausen codices, both old and new, were included. The picture cues *women in lab* and *ship captain* were widely used by McClelland and associates (reproduced in Smith, 1992), *director* and *man at desk* are pictures B and E from Heckhausen's original six-image assessment

battery (Schultheiss, 2001), while *soccer duel* has been widely used in research (e.g., Schultheiss & Rohde, 2002). Of the remaining four images, *chemist* and *gymnast* were utilized by Pang et al. (2009), while *piano lesson* has been pretested for its ability to arouse achievement imagery (Pang, 2010a). Reproductions of these images can be found in Appendix C.

*6.1.3.3. Measures of *sanAch*.*

The Personality Research Form (PRF; Jackson, 1974) has been used extensively by researchers to investigate the relationship between implicit and explicit motives in the big three domains of achievement, affiliation, and power (e.g., Koestner et al., 1988). Additionally, the Hope of Success and Fear of Failure scale (Schultheiss & Murray, 2002), which was developed to provide an explicit measure of HS and FF that was content-matched to categories present in the Heckhausen (1963b) coding system, has recently been employed by several researchers investigating implicit-explicit motive congruence (e.g., Thrash, Elliot, & Schultheiss, 2007). Both measures were administered immediately after completion of the comparison PSE in session 4.

6.1.4. Validation tasks.

6.1.4.1. Wisconsin card sorting test (WCST).

An adapted version of the WCST was administered during session three in order to assess and compare the predictive validity of the ambiguous, unambiguous, and comparison picture sets. Originally developed as a measure of executive function, it was reasoned that the challenging and dynamic nature of the WCST, in which subjects are required to constantly adapt their strategy in the face of changing card-sorting criteria,

makes it suitable for the arousal of the achievement motive and for the expression of achievement relevant behaviors. This reasoning is in line with Brunstein and Schmitt's (2004) argument that tests of concentration are well-suited to the assessment of motivationally relevant behaviors, since effective performance (operationalized both in terms of accuracy and speed) requires a great deal of mental effort.

The present adaptation of the WCST comprised three blocks, each consisting of a maximum of 128 trials. Participants were instructed to sort the cards appearing at the top of the screen into one of four piles, and were informed that they would receive immediate trial-by-trial feedback as to whether they had made the right or wrong choice. They were also advised that their performance would be judged both on the speed and the accuracy of their responses, and that they would receive a cash payment that reflected how well they performed. While extrinsic incentives such as financial reward have often been associated with explicit achievement motivation, the present task offered only a modest performance-related reward in order to ensure a base level of engagement with the task. The private administration setting and provision of self-referenced, rather than norm-referenced feedback, ensured that implicit rather than explicit achievement motivation would be the dominant motivational force aroused by the task.

The card stimuli in the WCST could be classified according to three independent criteria—color (red, blue, yellow, green), form (circle, star, cross, triangle), and number (one, two, three, four)—although participants were not given any information regarding which of these sorting criteria was being applied during each trial. Participants were required to deduce the nature of these changing expectations from the feedback provided

and to modify their responses accordingly. The WCST was programmed such that the sorting criterion being applied (e.g., color) remained constant for only four trials before switching to the next criterion (form, and subsequently, number). This four-trial cyclical redefinition of sorting rules continued until the participant had correctly identified six sorting criteria or all of the 128 cards were sorted, at which point the block would terminate. The first block was envisaged as a practice block to familiarize the participants with the task, while the second and third blocks were experimental blocks that yielded actual monetary payment. After the second block, participants were provided with a feedback summary of their performance and invited to try and improve their scores in the third and final block. The highest scoring participant earned \$4.94 SGD (Singapore Dollars) over the course of the three experimental blocks, while the lowest scoring participant earned \$0.88 SGD. Since self-referenced feedback has previously been shown to be highly effective in arousing implicit achievement motivation (Pang, 2010b), responses from the third block provided the accuracy (percentage of correct trials) and speed (mean trial latency) data that were to be predicted by HS.

Two criterion variables were derived from the adapted WCST. The percentage of correct responses over the course of the third block provided a measure of accuracy, while the mean latency in milliseconds (time elapsed between presentation of the card to be categorized and decision by participant) provided a measure of speed. As per the hypotheses outlined in section 4.1.4, it was anticipated that HS would positively predict sorting speed and negatively predict sorting accuracy. Such observations would be consistent with the HS-motivated individual's "greater tolerance for frustrations"

(Schultheiss & Brunstein, 2005, p. 11), which should manifest as a willingness to make a few categorization errors if pursuit of the overall goal of fast and accurate performance in the WCST is not compromised.

6.1.4.2. Balloon analogue risk task (BART).

In order to further assess the predictive validity of the ambiguous, unambiguous, and comparison picture sets, participants also completed a version of the BART. The BART assesses risk-taking behavior, and task scores have been found to be associated both with real-world risk-taking behaviors such as smoking (Lejuez et al., 2002) and with self-report measures of impulsivity and illicit drug use (Lejuez et al., 2003).

In the BART, participants are required to inflate a series of simulated balloons in order to earn money. The balloon is inflated by pressing an on-screen “pump” button, and money is accumulated with each successive pump until the balloon bursts, at which point the participant loses all the money gained on that trial. Participants can however choose to “bank” the money at any time and proceed to the next balloon in the series. Thus, with every pump the participant must balance the potential gain of accruing more money with the potential risk of bursting the balloon and losing everything they had gained on that particular trial. As in the adapted WCST, participants were paid real money in accordance with how they performed over the course of the task. The highest scoring participant received \$10.90 SGD while the lowest scoring participant received \$1.45 SGD. As in the WCST, the small monetary incentive was included only to ensure a base level of engagement with the task, and was not of a sufficient magnitude to arouse explicit motivation.

Two dependent measures were derived from the BART. Following the example of the Lejuez et al. (2003), the risk-taking measure was the average number of pumps on balloons that do not burst, with higher scores indicating greater risk-taking behavior. The second variable derived from BART was mean trial latency, which provided a measure of performance speed in a time-independent task (unlike the WCST, in which performance was explicitly related to speed). As per the hypotheses outlined in section 4.1.4, it was hypothesized that FF would negatively predict risk-taking, operationalized as the number of pumps on non-bursting trials, since the avoidant nature FF should manifest as a tendency to pump less in order to avoid bursting the balloons (i.e., failure). It was also anticipated that FF would negatively predict speed in the BART, since the time-irrelevant nature of the task should allow FF-motivated individuals to behave in their naturally more cautious and deliberative manner.

6.2. Results.

6.2.1. Test-retest reliability.

Simple bivariate correlations were conducted in order to establish the test-retest reliabilities of the new picture set, and to compare the reliability of the ambiguous and unambiguous variants (see tables 2 and 3). HS and FF scores derived from the session one PSE administration were correlated with those derived from the session two PSE administration in order to assess their stability over time.

Session one HS scores were found to be significantly correlated with session two HS scores in both the ambiguous and unambiguous conditions, indicating that both

ambiguous and unambiguous variants exhibit some degree of test-retest reliability.

However in the case of FF, a significant correlation between session one and session two motive scores was observed only in the ambiguous condition, and not in the unambiguous condition.

These results offer only partial support for H1.1, since the correlation coefficients—especially those for FF—fall short of the $r = .60$ value reported by Schultheiss and Pang (2007) in their meta-analysis of PSE studies. The results do however offer support for H2.2, suggesting that the ambiguous picture set exhibits better test-reliability than the less ambiguous picture set. In summary, the significant correlations between session one and session two HS and FF scores in the ambiguous condition suggest that the ambiguous variant of the new picture set exhibit test-retest reliability to a limited degree.

Table 2
Descriptive statistics and intercorrelations in the ambiguous condition

	S1 HS	S1 FF	S2 HS	S2 FF	S4 HS	S4 FF	saHS	saFF	sanAch
S1 HS									
S1 FF	.054								
S2 HS	.465**	.125							
S2 FF	-.150	.328*	.245						
S4 HS	.364*	.106	.778**	.322*					
S4 FF	-.167	.391*	-.071	.372*	.105				
saHS	-.143	.043	-.014	-.089	-.067	.088			
saFF	-.091	-.013	.119	-.028	.246	-.113	.184		
sanAch	.049	.124	-.001	-.147	-.152	-.066	-.104	.283	
M	12.99	6.27	13.77	7.19	17.79	6.00	3.52	3.28	9.49
SD	10.97	5.47	9.64	7.12	12.89	5.99	.34	.28	2.67

Note. S1 = session 1, S2, = session 2, S4 = session 4, saHS = explicit HS, saFF = explicit FF, sanAch = explicit nAch, ** $p < .01$ (2-tailed), * $p < .05$ (2-tailed).

Table 3
Descriptive statistics and intercorrelations in the unambiguous condition

	S1 HS	S1 FF	S2 HS	S2 FF	S4 HS	S4 FF	saHS	saFF	sanAch
S1 HS									
S1 FF	.205								
S2 HS	.390*	-.111							
S2 FF	-.128	.154	-.121						
S4 HS	.378*	-.036	.722**	-.079					
S4 FF	-.247	.001	-.209	.058	-.019				
saHS	.107	.091	.012	-.257	-.103	-.270			
saFF	.049	.067	.167	.224	.091	.300	-.004		
sanAch	-.327	-.170	-.091	.060	-.074	-.005	-.134	-.296	
<i>M</i>	22.99	10.57	22.69	11.03	23.76	6.24	3.54	3.26	.60
<i>SD</i>	16.73	4.98	12.80	6.07	14.35	5.25	.26	.25	.16

Note. S1 = session 1, S2, = session 2, S4 = session 4, saHS = explicit HS, saFF = explicit FF, sanAch = explicit nAch, ** $p < .01$ (2-tailed), * $p < .05$ (2-tailed).

Table 4

Descriptive statistics for validation task variables in studies 1B and 2C

		<i>M</i>	<i>SD</i>
WCST Speed (trial latency in ms)	Practice block	1737.38	342.59
	Post-feedback block	1372.94	399.43
WCST Accuracy (percentage of correct responses)	Practice block	50.71	13.16
	Post-feedback block	56.82	17.52
BART Speed (trial latency in ms)		444.49	296.48
BART Risk-Taking (number of pumps on non-bursting trials)		20.67	13.43

6.2.3. Predictive validity.

6.2.3.1. WCST.

Several hierarchical multiple linear regressions were conducted in order to assess the predictive validity of HS scores derived from the ambiguous, unambiguous, and comparison picture sets with respect to speed and accuracy in the WCST. Descriptive statistics for the dependent variables derived from the WCST can be found in table 4. In the first such regression analysis, post-feedback speed (mean response latency in milliseconds for the third experimental block) was predicted by practice block speed and the average of the HS scores from session one and session two. This analysis allowed for the predictive power of HS with respect to speed in the post-feedback trials to be assessed

while controlling for individual differences in task ability, as indicated by their performance on the practice trials. Practice trial latency (i.e., speed) was included as a predictor in the first step while the HS motive score was added in the second step. Practice trial latency was found to be a significant predictor of post-feedback trial latency ($\beta = .58$, $t(39) = 4.44$, $p < .001$), while ambiguous HS was also a significant predictor of post-feedback trial latency ($\beta = -.37$, $t(38) = -3.11$, $p < .01$). This result suggests that, in the ambiguous condition, individuals scoring higher for HS progressed through the task more quickly than those scoring lower for HS. Identical regression analyses revealed that neither unambiguous HS ($\beta = -.06$, $t(30) = -.36$, $p = .72$) nor comparison HS ($\beta = -.13$, $t(69) = -1.34$, $p = .18$) were significant predictors of post-feedback trial latency. H3.1 was therefore supported.

In the second round of regression analyses, post-feedback accuracy (percentage of correct responses in the third experimental block) was predicted by practice block accuracy and HS. Once again, this analytic approach allowed for the assessment of the predictive power of HS while controlling for baseline task ability, this time in terms of sorting accuracy, by including practice trial performance as the sole predictor in the first step before adding HS in the second step. Practice trial accuracy was found to be a significant predictor of post-feedback trial accuracy ($\beta = .37$, $t(39) = 2.46$, $p < .05$), while ambiguous HS was also a significant predictor of post-feedback trial accuracy ($\beta = -.36$, $t(38) = -2.54$, $p < .05$). This result indicates that, in the ambiguous condition, individuals scoring higher for HS were less accurate than those scoring lower for HS. Identical regression analyses revealed that neither unambiguous HS ($\beta = .12$, $t(30) = .76$, $p = .45$)

nor comparison HS ($\beta = .03$, $t(69) = -.27$, $p = .78$) were significant predictors of post-feedback trial accuracy, meaning that H3.2 was also supported.

6.2.3.2. BART.

Linear regression analyses were conducted to assess the predictive validity of FF scores derived from the ambiguous, unambiguous, and comparison picture sets in the context of the BART. Specifically, average session one and session two FF scores were used to predict risk-taking behavior, operationalized as the average number of pumps on non-bursting trials, as well as task speed, operationalized as mean trial latency. Descriptive statistics for the dependent variables derived from the BART can be found in table 4.

Risk-taking was not significantly predicted by ambiguous, unambiguous, or comparison FF. Contrary to the prediction in H4.1, these results implied that fear of failure might be unrelated to risk-taking behavior as measured by the BART, and that general risk-taking behaviors may unsuitable for use as outcome variables in FF predictive validity comparisons (although later results in study 2, see section 11.2.4.2, suggested otherwise). Mean trial latency however was significantly predicted by ambiguous FF ($\beta = .32$, $t(36) = 2.03$, $p = .05$), suggesting that individuals in the ambiguous condition who scored higher for FF spent longer on the task than those individuals in the same condition who scored lower for FF. Mean trial latency was not significantly predicted by either unambiguous FF ($\beta = -.03$, $t(27) = -.13$, $p = .90$) or comparison FF ($\beta = -.07$, $t(63) = -.54$, $p = .59$). H4.2 was therefore supported by the data.

These results indicate a substantial difference in the predictive validities of the ambiguous and unambiguous picture sets. This is remarkable given that they are identical save for two pictures. To further investigate the differences between these two picture set variants, participant motive scores were recalculated using only the six images that the two picture sets had in common (*basketball, blueprints, lecture theater, operating theater, skaters, and student*). The results of the HS analyses remained unchanged. Ambiguous HS was found to be a significant predictor of both post-feedback trial latency ($\beta = -.35, t(38) = -2.91, p < .01$) and post-feedback accuracy ($\beta = -.30, t(38) = -2.06, p < .05$), while unambiguous HS was a non-significant predictor in both cases. However, ambiguous FF derived from the six picture set no longer significantly predicted BART latency ($p = .14$).

6.3. Discussion.

The results of study 1B offer support for the reliability of the ambiguous variant of the novel picture set, and tentative support for its validity. The ambiguous picture set was found to exhibit a limited degree of test-retest reliability and satisfactory divergent validity. The ambiguous variant also outperformed the unambiguous version in all indices of psychometric fitness, and was more accurate in its prediction of theoretically relevant behaviors than the comparison picture set. These findings suggest that the ambiguous novel picture set is a better measure of implicit achievement motivation than a set of commonly used images, and the implications of these findings will be further elaborated in the general discussion.

However, while HS scores derived from the ambiguous variant significantly predicted both theoretically-relevant behaviors—speed and accuracy in a time-dependent

task—only one of the predictions regarding FF was borne out by the data. Ambiguous FF scores were found to predict speed in a time-independent task, but did not significantly predict risk-taking in the same task. Furthermore, comparison of the predictive validity of scores derived from the six images shared by the ambiguous and unambiguous variants, revealed that when this alternative strategy was adopted, only the predictions regarding HS were born out by the data. This latter finding suggests that while the addition of two non-achievement images improves the reliability and validity of an otherwise unambiguous picture set, the two non-achievement images also contribute to the validity of the resulting ambiguous picture set.

STUDY 2

7. Study 2: Overview

The aim of study 2 was to develop and subsequently validate a consolidated content coding system for the assessment of implicit achievement motivation: a system that retains the best and most useful features of the four systems described in the introduction, while simplifying administration as much as possible. In combination with the new picture set developed in study 1, this new coding system would represent an improved instrument for the assessment of implicit achievement motivation.

It was envisaged that this consolidated system would improve PSE assessment of *nAch*, HS, and FF in the four ways described in section 3.10:

- a) Incorporation of the HS/FF distinction
- b) Characterization of *nAch*, HS, and FF
- c) Simplification of administration
- d) Avoidance of motive contamination

Study 2 comprised three phases, each of which involved secondary analysis of data collected during study 1. In study 2A, the structure for a new consolidated coding system was developed using the data reduction technique of cluster analysis. Stories written in response to the set of commonly used images utilized in studies 1A and 1B were recoded for implicit achievement motivation using the McClelland et al. (1953), Winter (1991), and Pang (2006) systems, supplementing the scores obtained using the Heckhausen (1963b) system during study 1. Adopting the picture or story (rather than the individual) as the unit of analysis, these motive and subcategory scores were then clustered on the basis of correlational similarity, producing a dendrogram (a visual

representation of the association between motives and categories) that would subsequently direct the construction of a consolidated manual in the study 2B. In study 2C, the data from study 1B were recoded again using the nascent coding system. This allowed for the assessment of the reliability and validity of the new system using the same analytic techniques that established the psychometric fitness of the new ambiguous picture set in study 1.

The decision to use cluster analysis to combine existing coding systems, rather than building a new system from scratch, was borne of a desire to facilitate integration within the field of implicit achievement motivation, while at the same time characterizing these various motivational constructs, elucidating the relationships between them, and identifying instances of redundancy or motive contamination. As a data reduction technique that groups cases or variables on the basis of similarity, cluster analysis is well-suited to each of these purposes, since each involves the identification of similarities and differences between the various motives and their constituent subcategories. The specific utility of cluster analysis to each of these applications is discussed in the following paragraphs, while a more detailed discussion of the procedural mechanics can be found in section 9.1.4.

As outlined in the introduction, the use of multiple coding systems with different operational definitions of the motives they assess has created a somewhat fragmented field, where the applicability (or inapplicability) of findings from one tradition has not been rigorously examined. For example, HS as measured by the Heckhausen (1963b) and Pang (2006) systems is widely thought to be analogous to the unitary *nAch* construct measured by the McClelland et al. (1953) and Winter (1991) systems. Many researchers

have noted the strong correlations between McClelland et al.'s (1953) *nAch* and Heckhausen's (1963b) HS (e.g., Brunstein & Heckhausen, 2008), while Thrash and Hurst (2008, p. 222) state that "McClelland's *nAch* and Heckhausen's HS are so highly correlated that they are generally regarded as alternative indicators of the same approach-based motive." Despite these assertions, even cursory examinations of the coding categories that comprise these two systems reveal significant differences between *nAch* and HS. McClelland et al.'s (1953) *nAch* is calculated by summing both approach- and avoidance-related subcategories, and so may constitute an impure measure of approach-related achievement motivation, even though it correlates strongly with Heckhausen's (1963b) HS. Similarly, one of Winter's (1991) five criteria for scoring achievement motivation is affective concern over "failure, doing badly, or a lack of excellence", which pertains to outcomes that achievement-motivated individuals strive to avoid. This definitional uncertainty is borne of the fact that the compatibility of the various systems and their resulting motive constructs and has never been investigated at the subcategory level. Study 2A aimed to address these issues, providing the first empirical investigation of the relationships between four commonly used PSE coding systems for the assessment of the achievement motive. By using cluster analysis, study 2A investigated not only the correlational similarity of the superordinate motives derived from the various systems, but also the correlational similarities (and differences) between the subcategories that define them. This approach allowed for the construction of a novel system that yields information that can easily be integrated with findings from other research traditions.

With respect to point (b) and (c), studies 2A and 2B were designed to produce a nascent coding system that incorporated Heckhausen's distinction between HS and FF (a

critical property that is missing from today's most commonly utilized measure of *nAch*, the Winter system), while simplifying assessment to the greatest possible degree. Although the Heckhausen (1963b) system has been acknowledged as "the most advanced, systematic, comprehensive and theory-driven need Achievement scoring system available" (Schultheiss, 2001, p. 1), the majority of researchers prefer to utilize other systems, particularly Winter's (1994) system, in their achievement motive research. This preference for Winter's system can safely be attributed to two factors. Firstly, the Winter system allows for simultaneous assessment of *nAch*, *nAff*, and *nPow*, meaning that researchers interested in two or more motives can save substantial time by coding their data using a single multipurpose system. Secondly, the Winter system was designed in such a way as to simplify the scoring process, with subcategories being removed and definitions of achievement imagery simplified down to their basic essence. The use of cluster analysis to derive a new system in study 2A was intended to replicate the simplicity of Winter's system while incorporating the distinction between HS and FF. The use of cluster analysis allows for a two-pronged approach to enhancing simplicity, whereby issues of excessive similarity between subcategories (i.e., redundancy) and weak association of subcategories with parent motives can easily be identified by way of correlational (dis)similarity; an approach that was expected to yield improvements in validity, since weakly associated subcategories would no longer be allowed to influence the final motive scores. The results of the cluster analysis were also anticipated to support the validity of the HS/FF distinction by yielding a two-cluster solution.

Finally, the use of cluster analysis was also intended to address the issue of motive contamination identified in the introduction. By clustering achievement motive

subcategories along with Winter's *nAff* and *nPow* motive scores, cluster analysis allowed for quick and accurate identification of any subcategories that associate more closely with other motives than with *nAch*, HS, or FF. One example would be McClelland et al.'s subcategory of Nup, which captures instances of benevolent assistance towards achievement goal pursuit, and was expected to relate more strongly to *nAff* than to *nAch*. Since cluster analysis can unambiguously identify such instances of motive contamination, it would provide a stronger rationale for the elimination of such categories and for the development of a leaner, more valid system.

8. Study 2A: Developing the Coding System

8.1. Method.

8.1.1. Participants.

Two hundred and three undergraduates from a large Singapore university participated in study 2A in return for partial course credit. The sample comprised the participants who contributed data to study 1B, as well as many participants who contributed to the various pretesting phases in study 1A. The sample was 69.7% female, with a mean age of 21.55 years ($SD = 1.73$). In terms of ethnicity, the sample was 86.1% ethnically Chinese, 3.5% Malay, 4.5% ethnically Indian, and 1.5% Caucasian. The remaining 4.5% listed their ethnicity as "other". Informed consent was obtained in all cases, and participants were fully debriefed at the end of the experiment.

8.1.2. Design and procedure.

All participants in study completed a single eight-image PSE, as well as other tasks that were specific to the study they completed (see sections 5.1.2. and 6.1.2.). As in study 1, all components were administered online with no requirement for the participants

to attend laboratory sessions, and PSE administration followed the same format as in study 1. Image presentation order was randomized, and demographic data were collected either on the day of PSE administration or during an associated experimental session, depending on study allocation.

All stories were coded for *nAch*, HS, and FF using the McClelland et al. (1953), Winter (1994), and Pang (2006) systems. Heckhausen (1963b) scores obtained during study 1 were also used in the cluster analysis. In study 2, coding was carried out by a single rater who previously achieved >85% reliability with the practice materials associated with each of the four systems. In order to guard against idiosyncratic application of the various systems' coding rules, 10% of the stories were also scored for the various systems by a second coder. Inter-rater reliability was calculated for each of the four systems, and was found to exceed 85% in each case, indicating that the single coder was applying the rules of the various coding systems judiciously and without bias. Inter-rater reliability, operationalized as the inter-rater concordance rate, was calculated using the procedure outlined in Pang (2010b). The concordance rates for the McClelland et al. (1953), Heckhausen (1963b), Winter (1994), and Pang (2006) systems were .90, .99, .92, and .92 respectively.

8.1.3. Materials.

The pictures used were the comparison images from study 1B—*women in lab*, *ship captain*, *director*, *man at desk*, *soccer duel*, *piano lesson*, *chemist*, and *gymnast*—which were also administered to many participants in study 1A, as outlined in section

6.1.3.2. These pictures were chosen to ensure representation of all the major picture selection traditions.

8.1.4. Analysis.

8.1.4.1 Rationale and approach.

As outlined in the preceding section, the motive and subcategory scores from the four systems were analyzed using cluster analysis. Defined as “the art of finding groups in data” (Kaufman & Rousseeuw, 2005, p. 1), cluster analysis is a family of related statistical procedures that group objects on the basis of similarity, allowing for the identification of subgroups in data that may be of explanatory value. While cluster analytic techniques are typically used to group individuals or cases, such procedures can also be used to group variables, and it was to this end that cluster analysis was employed in the present study. One of the chief strengths of cluster analysis is its flexibility, and its freedom from the more restrictive assumptions that accompany inferential statistical techniques typically applied to similar problems. However, cluster analysis’ status as a descriptive, noninferential technique means that all cluster analyses are inherently exploratory, and the solutions or structures they reveal must be validated using other methods. This was the purpose of study 2C, which provides evidence for the validity of scores derived from the new coding system developed in study 2B, using the clustering solutions obtained in study 2A.

The central challenge of study 2A was to identify patterns of similarity and dissimilarity between a large number of variables: the motive and subcategory scores derived from the various content coding systems. In order to inform the development of a

consolidated coding system, the analysis was required to identify variables that are highly similar (such as coding subcategories with similar definitions that might be redundant) as well as variables that are highly dissimilar (such as coding categories that are more closely related to alternate motives, and so do not associate strongly with *nAch*, HS, or FF). When faced such a challenge—one of simplifying relationships between large numbers of variables and identifying larger, superordinate variables that explain these relationships—researchers would typically employ exploratory factor analysis. Those coding categories that load strongly onto a latent variable corresponding to either *nAch*, HS, or FF would be retained as good “items” for the assessment of that motive, while those that load weakly, or load strongly onto an alternate, non achievement-related latent construct, would be discarded.

Nonetheless, cluster analysis was preferred over factor analysis primarily on the grounds of informational richness, since the use of a hierarchical, agglomerative cluster analytic technique lends itself to the exploration of inherently hierarchical data. While it is possible that factor analysis may have yielded several latent variables corresponding to *nAch*, HS, or FF, the rotated factor loadings would have only indicated the relatedness of these factors to these latent variables, and therefore allows for only two levels of explanation. Interpretation of the dendrogram resulting from a hierarchical, agglomerative cluster analysis on the other hand, allows for the appraisal of a range of clustering solutions at different levels of the hierarchy, some of which speak of category redundancy (e.g., early clustering of two highly similar categories), while others provide insight into underlying motive structure (e.g., the formation of a two vs. three cluster solution after application of the stopping rule). As such, cluster analysis provides a greater richness of

information than factor analysis, and was therefore preferred.

As a diverse family of related techniques, cluster analysis can be approached in many different ways. This versatility requires researchers to make several important decisions, based on the nature of their data and the research question under investigation, when planning a cluster analysis. The most important decisions, which have received the most attention in the cluster analytic literature, concern the definition of similarity, the clustering method, and the stopping rule to be employed. However certain preliminary decisions, such as the unit of analysis, the conceptualization of the objects or variables to be clustered, and the decision whether to rescale or standardize the data, can also have important repercussions. The following sections provide an overview of these preliminary decisions before moving on to an in-depth discussion of similarity measures, clustering methods, and stopping rules.

8.1.4.2. Preliminary decisions.

In the present study, PSE stories, rather than participants, were chosen as the unit of analysis. This can be justified on both theoretical and pragmatic grounds. Firstly, choosing the story as the unit of analysis is more theoretically sensible given that the study's primary aim is to ascertain how coding categories and resulting motive scores co-occur in written narratives. The co-occurrence of two categories (e.g., Heckhausen's ES and McClelland et al.'s Ga+) in a single story is more meaningful than their co-occurrence in an aggregated set of eight stories, since they likely correspond to the same passage of text or part of the story. If categories or motives tend to cluster together within stories, this can be interpreted as them pertaining to similar verbal manifestations of motive imagery. Clustering of categories and motives within a person is not so readily interpreted.

Secondly, choosing the story as the unit of analysis yields a total sample size of $N = 1,384$, compared to a sample size of less than 200 when choosing the individual as the unit of analysis. While cluster analysis is more robust to small sample sizes than methods like factor analysis, many researchers still recommend large sample sizes on the grounds that they increase the likelihood that small clusters will be detected and not dismissed as outliers (Hair & Anderson, 2009).

It is also important to identify the types of variables to be clustered. While PSE-derived motive scores are typically assumed to be continuous variables (hence their frequent employment as predictors in linear regression analyses), this assumption proves to be much more tenuous when analyzing the output of individual PSE stories. Motive scores on specific stories are additive combinations of subcategory scores, which are best conceptualized as either binary or counts variables, depending on the coding system in question. The specific rules associated with each coding system dictate how frequently a given subcategory can be scored within a single story, and so determine the quality of the associated variable. For example, the Heckhausen (1963b) system specifies that a given subcategory (e.g., IS) can only be scored once per story, giving rise to a binary variable indicating the presence or absence of instrumental activity in pursuit of a positively framed achievement goal, whereas the Pang (2006) system places no such restrictions on the frequency of subcategory scoring, resulting in count data. In the present research, McClelland et al. (1953) and Heckhausen (1963b) subcategories were binary variables, while Pang (2006) subcategories and combined motive scores from each of the four systems were counts, meaning that the entire dataset was a combination of mixed binary/counts data. This fact had important ramifications for the selection of similarity

measures, and will be discussed further in the following section.

Another preliminary consideration was whether or not to standardize the variables. Standardization or rescaling is frequently recommended on the grounds that most cluster analytic procedures cannot distinguish between differences in the scaling of variables and differences in magnitude of the values they assume. Continuous variables are typically standardized by conversion to Z scores, while mixed data are often standardized by conversion to binary, rescaling to a range of one, or by utilizing a distance measures that incorporates a standardization procedure, such as the Mahalanobis distance or Gower coefficient (Romesburg, 2004; Yaghini, 2010). For the present purpose, rescaling was deemed unnecessary. The adoption of a correlational rather than distance-based similarity measure (the rationale for which is outlined in more detail in the following section) meant that differences in scale would not adversely impact the analysis, and so standardization of mixed variables such as these would result in a loss of information without any appreciable benefits. Consequently, no rescaling or standardization procedures were employed.

8.1.4.3. Measures of similarity, clustering procedures, and stopping rules.

Measures of inter-object or inter-variable similarity fall into three distinct classes: correlation, distance, or association. Association-based measures of similarity are appropriate only for purely nonmetric (i.e., nominal or ordinal) data and so will not be discussed further. For mixed datasets including binary and count variables, distance or correlation measures are both appropriate and the decision to adopt one or other largely comes down to which offers the most useful definition of similarity given the research question under investigation. The principle advantage of distance measures such as the

Euclidean distance, squared Euclidean distance, Manhattan distance, and Mahalanobis distance is that they incorporate magnitude into their definition of similarity. For many clustering applications, such as the identification of social classes or strata, such a definition is appropriate, since large difference in the magnitudes of clustering variables such as income and education should likely place individuals in separate clusters. For other clustering applications however, magnitudes are of secondary importance to the patterns of association between variables: whether the variables increase and decrease in unison irrespective of their absolute values. In such instances, correlational measures of similarity such as Pearson's r are more appropriate.

For the present research, patterns of association were likely to be more informative. Differences in magnitudes (i.e., the values a subcategory or motive score can assume) are to a large extent dictated by the rules of the parent scoring system, and comparisons between binary variables, which can only indicate the presence or absence of motive imagery, and count variables, which indicate differences in amount, are not meaningful. Patterns of association on the other hand, should be more revealing of genuine similarities between subcategory and motive scores derived from the various systems. If two categories rise and fall in a similar fashion across the entire sample of stories, this suggests that these two categories correspond to the same elements in the text, and so are redundant. Similarly, if one category varies in a way that is completely unrelated to any of the other categories, then it is likely to be measuring an unrelated construct and should be discarded. As such, the most commonly used measure of correlational similarity, Pearson's r , was selected for use in the cluster analysis.

While the similarity measure dictates how the similarity between two cases or

variables is calculated, the clustering procedure determines how this similarity information is used to inform the generation of clusters. Several decisions are required when choosing a clustering method. The researcher must first decide whether to employ a hierarchical or a non-hierarchical technique, and, if a hierarchical method is preferred, must then decide whether to use a divisive or agglomerative clustering procedure. Furthermore, the researcher must choose which clustering algorithm or algorithms to use during cluster formation.

A hierarchical, agglomerative clustering procedure was employed in the present research. The decision to employ a hierarchical technique was a straightforward one, since hierarchical methods are more appropriate for exploratory investigations such as this. Non-hierarchical techniques, such as K-means clustering, involve prior specification of the number of clusters in the final solution, and so are more typically used to support hypothesized structures in data. Such *a priori* assumptions regarding the relationships between motives and subcategories may obscure unanticipated relationships, and yield a less accurate final clustering solution. Furthermore, the use of a hierarchical technique lends itself to the examination of inherently hierarchical data such as PSE subcategory and motive scores (see section 8.1.4.1).

Decisions regarding clustering algorithms are typically less straightforward as each has its own advantages and disadvantages, and researchers frequently recommend employing multiple algorithms and looking for similarity in the solutions they provide (Dugard, Todman, & Staines, 2009). Among the most commonly employed clustering algorithms are the single linkage (or nearest neighbor) method, the complete linkage (or furthest neighbor) method, the average linkage method, the centroid method, and Ward's

method. Of these five algorithms, the complete linkage method and two variants of the average linkage method were selected for use in the present cluster analysis. While the single linkage method is versatile and can form clusters with wide range of shapes, it tends to produce long, snake-like clusters that contain dissimilar cases or variables and are difficult to interpret. Researchers similarly criticize the centroid method on the grounds that it produces “messy and confusing results” (Hair & Anderson, 2010, p. 532), while others caution against this method on the grounds that clusters merged at later stages are frequently more dissimilar than those merged at earlier stages (Norušis, 2008). Ward's method, despite being one of the most versatile and widely used clustering algorithms, is not appropriate for use with correlational similarity measures (Rasmussen, 1992). As such, these three methods were deemed unsuitable for the present analysis, leaving the complete linkage and average linkage algorithms as the methods of choice. Each of these methods eliminates the chaining problem associated with the single linkage approach and produces clustering solutions that are not distorted by outliers (Hair & Anderson, 2010).

The final critical decision when planning a hierarchical cluster analysis concerns the clustering solution criterion or *stopping rule* to be applied. Since the dendrograms produced by an agglomerative, hierarchical clustering procedure depict a range of possible solutions, stopping rules are necessary to identify the best clustering solution: the one that is most representative of the underlying data structure (Clatworthy et al., 2005). One commonly used class of stopping rules examine changes in heterogeneity associated with the combination of clusters at different stages of the agglomeration schedule in an agglomerative cluster analysis (Hair & Anderson, 2010). More heterogeneous clusters contain more dissimilar cases or variables, and therefore provide a less compelling

account of the underlying data. As such, choosing clustering solutions that precede large increases in heterogeneity provides a good balance between concerns of parsimony (minimizing the number of clusters) and homogeneity (ensuring similarity among cluster members). The purpose of stopping rules based on changes in heterogeneity is to identify these peaks of heterogeneity change and to select the clusters preceding them as the final cluster solution. While other measures of heterogeneity change exist, such as the root mean square standard deviation (RMSSTD) and the pseudo F statistic, the simplest and most widely used of these stopping rules is the percentage change in heterogeneity, which identifies the clustering solution preceding the largest percentage increase in heterogeneity as the final solution. This was the stopping rule that was employed in the present analysis. However, more recently developed clustering procedures have moved beyond simple stopping rules to provide more inferential indices of cluster solution correctness. Some researchers (e.g., Kerr & Churchill, 2001; Suzuki & Shimodaira, 2006) have used bootstrapping techniques to estimate the stability of solutions derived from cluster analysis, deriving statistics that quantify how well a particular clustering solution is supported by the data. In light of these developments, the present research also incorporated bootstrap resampling to supplement the heterogeneity-based stopping rule and provide additional evidence the best clustering solution.

8.1.4.4. Statistical software.

Cluster analysis of the motive and subcategory data was conducted using SPSS version 20.0 (IBM Corp., 2011), R version 3.0.1 (R Core Team, 2013), and the R package *pvclust* version 1.2-2 (Suzuki & Shimodaira, 2011). Two cluster analyses were conducted. The first employed Pearson's r as the measure of similarity and the average linkage

clustering algorithm, while the second employed Pearson's r and used the complete linkage clustering algorithm. The number of bootstrap resamples was set at 10,000 for each analysis.

8.2 Results.

To summarize the previous section, motive and subcategory variables derived from the McClelland et al. (1953), Heckhausen (1963b), Winter (1994), and Pang (2006) systems were cluster analyzed on the basis of correlational similarity using the Pearson's r statistic. PSE stories, rather than participants, were chosen as the unit of analysis, and no rescaling or standardization procedures were employed. The clustering procedures were hierarchical and agglomerative, and two separate cluster analyses were conducted: one adopting the complete linkage algorithm and another employing the average linkage algorithm. Two stopping rules were applied in order to establish the best clustering solutions. The first was based on percentage changes in heterogeneity during clustering, while the second employed a bootstrap resampling procedure to identify stable and unstable clustering solutions. The dendrogram output for the average linkage and complete linkage analyses can be found in figures 3 and 4. The average linkage analysis gave rise to a two cluster solution, providing evidence for the validity of the HS/FF distinction, while the complete linkage analysis yielded a three cluster solution that tentatively supports the HS/FF distinction while casting doubt on the usefulness of many HS-related subcategories. Both approaches gave rise to solutions in which n Ach- and HS-related motives and subcategories clustered together, away from those corresponding to FF. This finding provides evidence that n Ach and HS are equivalent in PSE assessment.

Figure 4

Dendrogram displaying the results of the complete linkage cluster analysis

In the average linkage analysis, the largest percentage change in heterogeneity occurred at the final clustering stage, when the two McClelland et al. (1953) categories corresponding to non-achievement imagery, TI and UI, clustered with all the other motives and categories. While this tentatively suggested a one-cluster solution, setting aside the non-motive relevant categories and concentrating on the clustering of the motive-relevant variables was judged to be more informative. Ignoring TI and UI, a two-cluster solution preceding the second largest change in heterogeneity (75.99%) was preferred. This clustering solution is face valid, with most of the HS- and *n*Ach-related motives and categories in one cluster and most of the FF-related motives and categories in another. Early candidates for deletion were revealed, as several FF subcategories, including McClelland et al.'s (1953) Bw and Pang's (2006) LD and O, were found to associate with the *n*Ach/HS cluster. Similarly, the dendrogram revealed several categories that clustered very early in the agglomeration schedule, such as Heckhausen's (1963b) HS and ST, that were therefore candidates for consolidation. It is also noteworthy that in this solution, Winter's *n*Aff and *n*Pow were found in separate clusters, and that they merged with their final clusters relatively late in the agglomeration schedule.

In the complete linkage analysis, the largest percentage change in heterogeneity (235%) occurred when three clusters were joined. Of these three clusters, one contained only FF-related categories and motives, one contained a mixture of HS- and *n*Ach-related categories and motives, while the third contained a number of HS-related categories and Winter's (1991) *n*Aff and *n*Pow motives. In this solution, McClelland et al.'s (1953) TI and UI failed to cluster at all. While the interpretation of these results is less straightforward, the dendrogram also exhibits some face validity, with FF-related motives

and categories clustering away from those associated with *nAch* and HS. The categories that were identified as candidates for deletion in the average linkage analysis were also flagged here, as they were once again found to cluster away from the other FF categories. The three cluster solution derived from the complete linkage analysis also identified candidates for consolidation on the grounds of early clustering, such as the positive expectation categories from the McClelland et al. (1953), Heckhausen (1963b), and Pang (2006) systems.

While the percentage changes in heterogeneity indicated appropriate cluster solutions for the two analyses, examination of the bootstrap probabilities associated with various cluster solutions provided further insight. For every cluster in the agglomeration schedule, Suzuki and Shimodaira's (2011; 2006) *pvclust* package calculated two *p* values that provided estimates of cluster stability: the approximately unbiased (AU; indicated in red in figures 3 and 4) and bootstrap probability (BP; indicated in green in figures 3 and 4) values. Since the authors recommend using the AU *p* value on the grounds that it provides a better approximation to the unbiased *p* value, the AU estimations of cluster stability were the focus of the analysis. For a cluster with an AU *p* value greater than .95, the null hypothesis that the cluster does not exist is rejected at the .05 significance level, which suggests that the cluster represents a stable group within the data.

For the average linkage analysis, the AU values found in figure 3 did not strongly support the two cluster solution identified by the percentage change in heterogeneity stopping rule, since *p* values of .79 for the FF cluster and .77 for the *nAch*/HS cluster were not significant at the .05 level. However, several subclusters within these two clusters exhibited enhanced stability. A *nAch*/HS subcluster that omitted FF-related

subcategories and Winter's (1994) *nAff* motive (highlighted in figure 5) had an associated p value of .98, indicating a highly stable cluster. This solution was also theoretically appealing, since it eliminated subcategories and motive scores associated with alternate motives, leaving only *nAch*- and HS-related constructs. While none of the higher level FF-related subclusters had a p value exceeding .95, one theoretically sensible clustering solution (highlighted in figure 5) was significant at the $p = .1$ level. Critically, this cluster also omitted irrelevant constructs such as Winter's (1994) *nPow*, along with several FF-related subcategories that may be contaminated with power-related imagery, such as Pang's (2006) QCO and Heckhausen's (1963b) C.

For the complete linkage analysis, the three clusters identified by the heterogeneity-dependent stopping rule all exhibited relatively weak stability. The FF cluster had an AU p value of .66, while the *nAch*/HS and HS/*nAff*/*nPow* clusters had p values of .86 and .84 respectively. Similarly, the various subclusters contained within each cluster were also unstable, with the exception of a few early-stage clusters that were more informative regarding subcategory redundancy than higher-level motive structure. Given that the clustering solutions obtained in the average linkage analysis were both more stable and theoretically sensible than those obtained in the complete linkage analysis, the solution highlighted in figures 3 and 5 was chosen to guide the construction of the consolidated coding system.

Figure 5
Clustering solution derived from the average linkage analysis

8.3 Discussion.

The results of the average linkage cluster analysis yielded two important insights into the relationships between the various conceptualizations of the achievement motive. Firstly, the two-cluster solution attests to the validity of the HS/FF distinction, and reinforces the notion that the approach- and avoidance-related manifestations of achievement motivation are distinct and independent entities. While much evidence suggests HS and FF are independent motives with different outcomes (Pang, 2010a), many researchers ignore this distinction when conducting research into the achievement motive. It is therefore hoped that these results, which represent a novel methodological approach to testing the HS/FF distinction, reinforce the contrast between these two components of the achievement motive.

Secondly, the clustering solution also attests to the functional equivalence of *n*Ach, as measured by the McClelland et al. (1953) and Winter (1994) systems, and HS, as measured by the Heckhausen (1963b) and Pang (2006) systems. While many researchers have observed strong associations between *n*Ach and HS (Heckhausen, 1963a), their equivalence has not yet been investigated at the subcategory level. Previous observations that HS and *n*Ach correlate strongly while FF and *n*Ach correlate only weakly, if at all (Heckhausen, 1963a), suggest that McClelland et al.'s (1953) system is better thought of as a measure of approach-related achievement motivation, albeit one that is to some extent contaminated with avoidance-related subcategories such as Ga- and G-. The results of the present study are consistent with this account. The McClelland et al. (1953) and Winter (1994) *n*Ach motive categories cluster with the Heckhausen (1963b) and Pang (2006) HS categories in a 12-member cluster that exhibits very high stability ($AU\ p < .001$), while

the two FF motive scores associate in an entirely different cluster. This finding has important implications for research on the implicit achievement motive, as researchers will now have a stronger basis for comparing findings from the various assessment traditions, given the present demonstration of equivalence between *n*Ach and HS. Nonetheless, it is important to note that all of the negatively valenced McClelland et al. (1953) subcategories (Ga-, G-, and I-) clustered with FF-related subcategories from the Heckhausen (1963b) and Pang (2006) systems. This finding empirically confirms suspicions that the McClelland et al. (1953) system is either (a) an impure measure of approach-related achievement motivation, or (b) an imbalanced and conflated measure of approach and avoidance achievement motivation that does not adequately capture the motive to avoid failure. While the absence of subcategories meant that the Winter (1991) system could not be examined in the same way, it is highly likely that the system suffers from the same problems, given its derivation from the McClelland et al. (1953) system and its inclusion of failure-related imagery in the scoring criteria.

The results of the cluster analysis provided a wealth of additional information that could inform the development of the consolidated coding system. The dendrogram indicated the presence of two stable higher level clusters corresponding to *n*Ach/HS and FF, each containing the appropriate motive scores along with several of the most closely associated (and hence most representative) subcategories. These important subcategories were retained for incorporation into the consolidated system, while those that clustered outside the *n*Ach/HS and FF clusters were discarded. Furthermore, categories contained within the two motive clusters that were found to cluster extremely early in the agglomeration schedule were marked for consolidation into singular categories in the new

system. By acknowledging and incorporating the HS/FF distinction that was so apparent in the results of the cluster analysis, it was hoped that the new system would address the weaknesses of the McClelland et al. (1953) and Winter (1994) systems outlined previously. The specific categories that were selected or combined, as well as the approach to writing their corresponding scoring rules for the new system, are described in the next section.

9. Study 2B: Formalizing the Coding System

9.1. Defining the coding system.

The translation of the cluster analysis results into a functional coding system took place in two stages. Firstly, the dendrogram structure, along with the associated AU p values for various clustering solutions, identified those categories that were to be grouped under *nAch/HS* or *FF*, as well as those to be discarded. Secondly, this new motive/subcategory structure was used as a template to write the consolidated coding system. These two stages are described separately in the following sections.

9.1.1. *nAch/HS*.

In the *nAch/HS* cluster, the 12 members of the subcluster identified in section 8.3 were consolidated into a single category termed *need for success* (NS). Disregarding the variables corresponding to resultant motives (e.g., *nAch*, *HS*) or motive combinations (e.g., *NH*, *AM*), the categories to be grouped under NS were: McClelland et al.'s (1953) *AI*, *N*, and *AchTh*; Heckhausen's (1963b) *NS* and *ST*; and Pang's (2006) *PAG*. Of these categories, *N*, *NS*, and *PAG* all refer to an experienced achievement need and are highly similar in their definitions, and can therefore be consolidated into a single subcategory reflecting an experienced need to pursue a positively framed achievement goal. The 12-member cluster also contains two categories relating to a *nAch* or *HS* thema, which in their parent systems serve to boost their associated motive scores when the motive imagery is elaborated to such an extent that it becomes the central theme of the text or narrative. However, the mechanics of scoring, particularly in the Heckhausen (1963b) system, give rise to a situation in which thema is almost always scored as a direct result of scoring need, and the close clustering of these subcategories in the present results

reinforces the impression that the separate assessment of thema categories complicates scoring while contributing very little in terms of explanatory power. Consequently, the thema categories were discarded, meaning that the remaining AI, N, NS, and PAG provided the basis for constructing the definition associated with the need for success cluster.

Three other highly stable HS subclusters also possess AU p values greater than .99. These clusters are highly homogeneous, grouping categories with very similar definitions. The first, which contains the Ga+, ES, and ES subcategories from the McClelland et al. (1953), Heckhausen (1963b), and Pang (2006) systems respectively, was comprised entirely of subcategories that indicate the presence of expectations regarding positive achievement outcomes. This subcluster was to define an *expectation of success* subcategory in the nascent coding system. The second highly stable HS subcluster contained the subcategories S, from the Pang (2006) system, and I+, from the McClelland et al. (1953) system, both of which indicate the presence of positive achievement outcomes or successes. These two categories were consequently grouped to form a *success* cluster/subcategory. Thirdly, cluster analysis also revealed a highly stable *positive affect* subcluster, containing the G+ (McClelland et al., 1953), A+ (Heckhausen, 1963b), and E+ (Pang, 2006) subcategories, each of which is scored when an individual experiences positive affect during the course of an achievement task, usually as a result of success or mastery. The definitions associated with these subcategories would define a positive affect subcategory in the consolidated system.

The remaining two *nAch*/HS subclusters exhibit lower stability. One, which includes the two praise categories from the Heckhausen (1963b) and Pang (2006) systems,

has an AU p value of .96, which probably reflects the more all-encompassing definition of praise in the Pang (2006) system. The other subcluster has a p value of .63, and includes three categories pertaining to instrumental activity: I? from the McClelland et al. (1953) system, IS from the Heckhausen (1963b) system, and SR from the Pang (2006) system. Despite their lower stability, these two subclusters were tentatively chosen to be incorporated into the consolidated system. The first was logically termed *praise*, while the second one was termed *effortful activity*, a name that combines the Heckhausen (1963b) and McClelland et al. (1953) reference to instrumental activity with Pang's (2006) emphasis on additional effort and hard work.

Several categories that failed to group with the n Ach/HS subcluster were discarded. Pang's (2006) EE failed to cluster with any other subcategories, while her PER category clustered with two FF-related categories: McClelland et al.'s (1953) Bw and Pang's (2006) O. This latter finding confirms anecdotal evidence that motive images pertaining to persistence and obstacles often coexist owing to definitional overlap (Pang, personal communication, 2013), something that is problematic when each category is meant to relate to independent motive tendencies. Similarly, McClelland et al.'s (1953) Nup subcategory was found to cluster with Winter's (1994) n Aff motive, indicating that this subcategory relates more closely to the need for affiliation, vindicating Heckhausen's decision to drop the category when constructing his coding system. None of these categories were chosen for incorporation into the new system.

In total six HS subcategories were derived from the cluster analysis results: need for success, expectation of success, positive affect, praise, effortful activity, and success.

9.1.2. FF

Fewer subclusters emerged in the FF cluster, and they were generally less stable than their *n*Ach/HS counterparts. Heckhausen's (1963b) NF failed to cluster with any other subcategories within the FF cluster, although this is unsurprising given that it is the only negatively framed achievement need category in any of the systems. Despite this result, NF was chosen for inclusion as a subcategory given the prominence of need categories or criteria in all four systems, and the desire to create a system that was definitionally balanced with respect to HS and FF. A relatively stable negative affect subcluster did emerge, including McClelland et al.'s (1953) G- and Heckhausen's (1963b) A-. Given the highly similar definitions of these two subcategories, they were selected for grouping as a combined subcategory labeled A- in the new system. Similarly, the dendrogram also featured a somewhat stable ($p = .33$) FF subcluster containing the negative expectation categories from the various systems (Ga-, EF, and EF), which were also chosen for grouping an *expectation of failure* (EF) subcategory in the new system, largely on theoretical grounds. The outcome categories also clustered within the FF cluster and were selected for grouping as *failure* (F) in the new system. Unlike for HS, the thema categories clustered more closely with the failure outcome categories than the need category. This result is likely a direct consequence of the Heckhausen (1963b) scoring mechanics, since scoring F is one of the criteria for scoring FT, whereas there is no category for success outcome in the Heckhausen (1963b) system. Nonetheless, the same rationale regarding the thema category was applied and it was excluded from the new system on the grounds of redundancy.

Three other subcategories also clustered within the FF cluster: Pang's (2006) IE

and McClelland et al.'s (1953) I- and Bp. The latter two categories clustered together despite their very different definitions, while IE clustered with Pang's (2006) fear of failure motive category, and eventually with the negative affect subcategories.

Incorporation of these categories into the new system was guided largely by theory. Given that I- indicates the failure of an achievement related instrumental act, it was incorporated into the new system's failure category, since this mirrors the approach taken in the case of HS. Since Bp and IE both reflect the inhibition of achievement-related action by forces arising from within the person, they were grouped into a single subcategory labeled *inhibition*. While these strategies do not directly reflect the cluster analytic output, they do provide a simpler structure for the new coding system, and can therefore be justified on the grounds of parsimony.

Many FF subcategories were excluded from the combined system. Pang's (2006) O and LD, and McClelland et al.'s (1953) Bw were all discarded on the grounds that they clustered more closely with HS, and thus were more strongly associated with approach-related achievement motivation. Heckhausen's (1963b) C and IF were excluded because they failed to cluster with the main FF cluster or with any other FF-related subcategories, while Pang's (2006) QCO was excluded due to its clustering with Winter's (1994) *nPow*, which suggests contamination with the power motive.

In total five FF subcategories were derived from the cluster analysis results: need to avoid failure, expectation of failure, negative affect, inhibition, and failure.

9.2. Writing the coding system.

The final version of the new coding system is reproduced in full in appendix D.

Two key principles drove the authorship of the new system: streamlining the rules and structure while staying true to the definitions of the constituent categories. It was hoped that this approach would give rise to a system that exhibits the simplicity of the Winter (1994) system while incorporating the approach-avoidance distinction of Heckhausen (1963b) and Pang (2006).

To this end, several scoring conventions were adopted from the Winter (1994) system. First and foremost, Winter's decision to drop scoring at the subcategory level and simply score for the presence or absence of HS and FF imagery was adopted. In the Winter system, one scores *nAch* whenever one of five criteria are met, with the additional restriction that only one instance of *nAch* can be scored per sentence unless multiple instances are interspersed with imagery corresponding to another motive, in which case multiple instances can be scored. Consequently, the subcategories identified for inclusion in section 9.1 were reconceptualized as scoring *criteria* in the Winter (1994) sense: six in the case of HS and five in the case of FF. In the new system, HS and FF are scored whenever one of their associated criteria is met, with the restriction that HS and FF can only be scored once per sentence. Discussions with several coders trained in the McClelland et al. (1953) and Heckhausen (1963b) systems (e.g., Tan, personal communication, 2011) have revealed a shared impression that coding with both systems is made more time-consuming by constantly having to decide whether particular instances of motive imagery better match the definition of one subcategory versus another. For example, when encountering negative affect that accompanies participation in an achievement-related activity, one is frequently uncertain as to whether to score A- or EF using the Heckhausen (1963b) system, since it is often unclear whether the negative affect

arises from anticipated or actual failure. Abandoning subcategories in favor of scoring criteria resolves this issue, as once one of the criteria has been met there is no need to worry about whether others have also been satisfied.

Following the precedent set by Pang (2006), the Winter (1994) convention that individual types of achievement motive imagery can be scored multiple times in a given narrative was also adopted. Both the McClelland et al. (1953) and Heckhausen (1963b) systems stipulate that each subcategory can only be scored once per story, a questionable practice that artificially compresses the range of scores and potentially reduces their validity. The consolidated system allows for multiple scorings, with the only restriction being that HS and FF can each be scored only once per sentence. Scoring was also simplified yet further by removing Winter's restriction that adjacent sentences cannot be scored for the same motive unless separated by an instance of alternate motive imagery, as this is another practice that unnecessarily compresses the range of motive scores.

In the new system, no initial decision is required about whether a narrative is achievement-related or not. This approach follows the decision of Heckhausen (1963b), Winter (1994), Pang (2006) to drop this convention and simply assess each sentence individually for the presence of achievement imagery. Following McClelland et al. (1953), achievement imagery in the new system is defined as “the presence of a standard of excellence”, which can be either “personal (e.g., concern with improvement, mastery) or interpersonal (e.g., concern with competition, performance)”. These definitions are further qualified with the statement that, in order to qualify as achievement imagery, actions or internal states linked to a standard of excellence must be “self-endorsed and non-trivial”. These constitute the most fundamental rules of the new scoring system,

which are then further elaborated in the definitions for the various criteria, which are for the most part amalgamations of the definitions provided in each constituent system.

However in some cases, the criterion definitions have either been widened or narrowed in scope for ease and accuracy of scoring. For example, both the positive affect and negative affect criteria in the new system pertain only to emotional states that arise *as a result* of engagement in achievement related activity. The definition was narrowed to as to exclude excessive scoring of words such as “stressed”, “nervous”, or “excited” that often refer to unvalenced arousal yet cause significant problems when scoring using the Heckhausen (1963b) system. Conversely, and following the approach taken by Pang (2006), the definitions of praise and positive/negative expectation were widened to include statements made by the author or narrator of the verbal material, since these should be equally indicative of an underlying achievement need. The definition of effortful activity, which is in many ways similar to Heckhausen's (1963b) IS category, was narrowed to only encompass achievement-oriented instrumental acts that involve additional effort, in order to eliminate the problem of excessive scoring of IS associated with the Heckhausen (1963b) system (Pang, personal communication, 2012).

Several other features were also incorporated to make the system more user friendly and to enhance scoring accuracy. Each criterion is illustrated with three positive and three negative examples, indicating when motive imagery should and should not be scored. In the case of the negative examples, additional explanation is provided to illustrate why these phrases do not qualify. Furthermore, the practice stories (which the novice coder should score with at least 85% accuracy before moving on to scoring real data), are also elaborated with a rationale for the scoring decisions. Often novice coders

are left perplexed by these master scores that accompany the McClelland et al. (1953), Heckhausen (1963b), and Winter (1994) systems, since there is no explanation as to why certain phrases were scored and others were not. The inclusion of a scoring rationale was intended to solve this problem and to facilitate learning of the new system.

10. Study 2C: Validating and Refining the Coding System

As outlined in section 7, study 2C was intended to assess the psychometric properties of the nascent coding system constructed in study 2B. To this end, PSE data collected in study 1B was recoded using five different variants of the nascent consolidated system, with a view to selecting the best-performing of the five variants as the final consolidated system. Since the results of study 1 suggested that motive scores derived from the unambiguous and comparison picture sets are less valid than those derived from the unambiguous picture set, only those stories written by participants assigned to the ambiguous condition in study 1B were recoded. Correlation of motive scores obtained in sessions one and two once again provided an index of test-retest reliability, while the average motive scores from the first and second sessions were used to predict the same theoretically relevant behaviors in the WCST and BART, in order to provide measures of predictive validity.

Comparing five variants provided a more comprehensive evaluation of the coding system structure identified in studies 2A and 2B. The initial consolidated coding system described in study 2B constituted the first version (V1) of the new system. Each of the four additional variants (V2-V5) was a simplified version of V1. Given that several scoring criteria in the new system, such as EA and I, corresponded to less stable subclusters in the two-cluster HS/FF solution, it was reasoned that omission of these criteria may give rise to a coding system with improved psychometric properties, since the more stable, core criteria are stronger indicators of the underlying motive. If a simpler system with fewer criteria was found to enhance the predictive validity of the PSE assessment, then that system would be preferred.

The first variant included all the criteria identified in studies 2B and 2C. The second variant excluded the praise, effortful activity, and inhibition criteria, yielding a symmetrical system with four criteria each for HS (need for success, expectation of success, positive affect, success) and FF (need to avoid failure, expectation of failure, negative affect, failure). The third and fourth variants simplified the system yet further by excluding the affect and expectation criteria respectively, while the fifth variant included only the need and outcome categories.

Following the rationale of study 1B, the following hypotheses were made: Firstly, it was hypothesized that both HS and FF scores derived from the consolidated system would display a one-week stability coefficient of about .60 (H1), providing evidence of test-retest reliability. Secondly, it was hypothesized that the HS and FF scores derived from the consolidated system would be unrelated to participants' scores on two measures of explicit achievement motivation: Jackson's (1974) PRF (H2.1) and the Hope of Success and Fear of Failure Scale (Schultheiss & Murray, 2002) (H2.2), thus providing evidence of divergent validity. While these hypotheses mirror those made regarding the new ambiguous picture set in study 1B, here it was also hypothesized that the new system would display *convergent* validity. Since scores derived from a valid measure should correlate with those of another measure of the same construct, the validity of the consolidated system was further assessed by correlating the resulting HS and FF scores with those obtained using the Heckhausen (1963b) system in study 1B. It was hypothesized that HS and FF scores from the consolidated system would correlate significantly with their counterpart scores from the Heckhausen (1963b) system (H3). Finally, as in study 1B, it was hypothesized that the consolidated system would exhibit

predictive validity. Specifically, it was hypothesized that HS scores from the consolidated system would positively predict sorting speed (H4.1) and negatively predict sorting accuracy (H4.2) in the WCST, while FF scores derived from the consolidated coding system would negatively predict both risk-taking (H5.1) and task speed (H5.2) in the BART.

10.1. Method.

As with the other parts of study 2, study 2C involved the recoding and reanalysis of data collected during the first study. Specifically, all participants in study 2C ($N = 41$) had previously contributed data to the ambiguous condition of study 1B. The sample was 56.4% female, with a mean age of 21.36 years ($SD = 1.83$). 76.9% participants were ethnically Chinese, 2.6% were Malay, 10.3% were ethnically Indian, while the remaining 10.6% listed their ethnicity as “other”. The design, procedure, and materials employed in study 2C are identical to those used in study 1B, and more details can be found in sections 6.1.2 to 6.1.4.

10.2. Results.

10.2.1. Test-retest reliability.

As in study 1B, simple bivariate correlations between session one and session two scores provided an indication of motive score stability over time. All motive scores were corrected for word count using the procedure described in section 6.1.2 prior to analysis. Correlations were calculated separately for HS and FF, and the results for the five different coding system variants can be found in table 5.

Session one scores were significantly correlated with session two scores at the $p <$

.05 level in all cases, with one notable exception. Session one and session two HS scores derived from the first variant were only marginally correlated ($p = .073$), with a low stability coefficient of $r = .286$. Generally the temporal stabilities of the FF scores were found to be greater than that of the HS scores, a reversal of the pattern found in study 1B. Nonetheless, each of the other coding system variants could be said to exhibit reasonable test-retest reliability, since all correlations were significant with r values in the .35 to .47 range. This result compares favorably with the test-retest reliability of the Heckhausen system scores found in table 2, and actually represents a marked improvement in the stability of the FF scores.

These results offer partial support for H1. While the observed stability coefficients fell short of the .60 level observed by Schultheiss and Pang (2007), first and second session scores were significantly correlated in all cases, with the exception of HS scores derived using version one. Furthermore, the reliability statistics are comparable to those observed for the Heckhausen (1963b) system in study 1. As such, coding system variants two, three, four, and five can be said to exhibit a limited degree of test-retest reliability (H1 partially supported).

10.2.2. Divergent validity.

Scores derived from the five system variants were correlated with explicit measures of n Ach, HS, and FF in order to investigate divergent validity of the consolidated coding system. For each of the five variants, HS and FF scores were correlated both with scores on the Jackson (1974) PRF and with their explicit counterpart scores from the Schultheiss and Murray (2002) Hope of Success and Fear of Failure Scale. None of the HS and FF scores from any of the system versions correlated

significantly with the explicit measures. Hypotheses 2.1 and 2.2 were therefore supported by the data.

10.2.3. Convergent validity.

The average of the session one and two motive scores from the five systems were correlated with their corresponding Heckhausen (1963b) system averages from study 1B to provide an indication of convergent validity. Since scores derived from a valid measure should correlate significantly with those derived from a previously validated measure of the same construct, these statistics provided a means of assessing the validity of the new system. Bivariate correlations were calculated separately for HS and FF, and the results of the analysis can be found in table 5.

HS and FF scores from all systems were significantly correlated with their Heckhausen (1963b) counterparts at the $p < .01$ level, and so all system variants can be said to exhibit convergent validity. In the case of HS, r coefficients ranged between .812 and .766, with versions two and three exhibiting the strongest associations. Correlations for FF were more varied. Convergent validity was found to be highest for version two ($r = .743$) and lowest for version five ($r = .560$). Considering both the HS and FF results, version two exhibited the greatest convergent validity of the various coding systems. Hypothesis 3 was therefore supported by the data.

10.2.4. Predictive validity.

10.2.4.1. WCST.

As in study 1B, a series of hierarchical multiple linear regressions were conducted in order to assess the predictive validity of HS scores derived from the different variants of the new coding system. Descriptive statistics for the dependent variables derived from

the WCST can be found in table 6. In each case, the average of the word count corrected HS scores obtained from sessions one and two were used to predict post-feedback task performance after controlling for baseline task ability. It was hypothesized that HS would positively predict sorting speed and negatively predict sorting accuracy, and separate regression analyses were conducted for each coding system variant.

Table 5
Psychometric properties of the five coding system versions

			Coding System Version				
			V1	V2	V3	V4	V5
Test-Retest Reliability	HS	r	.286	.344	.373	.354	.389
		p	.073	.028	.016	.023	.012
	FF	r	.428	.435	.388	.470	.435
		p	.006	.004	.012	.002	.004
Convergent Validity	HS	r	.766	.803	.812	.773	.779
		p	< .001	< .001	< .001	< .001	< .001
	FF	r	.725	.743	.681	.660	.560
		p	< .001	< .001	< .001	< .001	< .001
Predictive Validity: WCST Speed	HS	β	-.256	-.302	-.287	-.314	-.295
		t	-2.043	-2.456	-2.313	-2.522	-2.356
		p	.048	.019	.026	.016	.024
Predictive Validity: WCST Accuracy	HS	β	-.316	-.306	-.307	-.341	-.340
		t	-2.163	-2.082	-2.094	-2.292	-2.293
		p	.037	.044	.043	.028	.027
Predictive Validity: BART Risk	FF	β	-.368	-.384	-.317	-.355	-.250
		t	-1.938	-2.040	-1.635	-1.863	-1.264
		p	.064	.052	.115	.075	.218

In the speed analyses, practice trial speed (block one) and HS (block two) found to significantly predict post-feedback speed in all cases, providing evidence for the predictive validity of all five systems. The β , t , and p values associated with the HS scores derived from each of the five variants can be found in table 5. HS derived from version four was the strongest predictor of post-feedback trial latency ($\beta = -.31$, $t(38) = -2.52$, $p < .05$), while HS from the first version was the weakest predictor ($\beta = -.26$, $t(38) = -2.04$, $p < .05$). All five HS versions were also significant predictors of sorting accuracy. In each multiple regression analysis, mean practice trial accuracy was found to be a significant positive predictor of post-feedback sorting accuracy, while HS was found to be a significant negative predictor. β , t , and p values for the five HS variants can be found in table 5. Version four HS was the strongest predictor ($\beta = -.34$, $t(38) = -2.29$, $p < .05$), while version two HS was the weakest predictor ($\beta = -.31$, $t(38) = -2.08$, $p < .05$), although all were significant. H4.1 and H4.2 were therefore supported.

10.2.4.2. BART.

Two sets of regression analyses were conducted to assess the predictive validity of FF scores derived from the five coding system variants. Descriptive statistics for the dependent variables derived from the BART can be found in table 6. In the first set of analyses, the average number of pumps on non-bursting trials, a measure of risk-taking, was regressed onto FF. Average session one and session two FF did not significantly predict risk-taking in all cases, although β values were of the correct valence (negative), and p values approached significance for versions one and two ($p = .16$). Similarly, the regression analyses of task speed also yielded non-significant results. In each case, FF was found to be a non-significant predictor of mean trial latency.

Despite these results, supplementary analyses offered tentative support for the predictive validity of FF scores derived from versions one, two, and four. Examination of the mean trial latencies revealed that many participants were responding very quickly throughout the BART, some spending as little as one tenth of a second contemplating each inflation decision. When the most rapid responders (mean trial latency < .3 seconds) were excluded from the analysis, a different pattern of risk-taking results emerged. Version two FF was found to significantly negatively predict the average number of pumps on non-bursting trials ($\beta = -.38$, $t(38) = -2.04$, $p = .052$), while FF derived from versions one ($p = .064$) and four ($p = .075$) were marginally significant predictors of risk-taking (see table 5 for more details). These results offer partial support for H5.2, although H5.1 was not supported by the data.

Table 6

Descriptive statistics and intercorrelations for motive scores in study 2C

		S1 <i>M</i>	S1 <i>SD</i>	S2 <i>M</i>	S2 <i>SD</i>	S1 <i>r</i> (HS*FF)	S2 <i>r</i> (HS*FF)
V1	HS	10.67	7.40	11.29	6.34	.115	.162
	FF	5.67	3.94	4.73	5.01		
V2	HS	7.71	6.84	7.28	5.35	.078	.168
	FF	5.42	3.75	4.41	4.83		
V3	HS	7.48	6.71	6.87	5.22	-.084	.086
	FF	4.04	2.84	3.15	3.96		
V4	HS	6.73	6.42	5.61	5.27	-.017	.109
	FF	4.15	3.31	3.48	4.16		
V5	HS	6.50	6.31	5.21	5.16	-.227	.002
	FF	2.77	2.42	2.22	3.25		

Note. S1 = session 1, S2, = session 2

10.3. Discussion.

Generally, the results of study 2C attest to the reliability and validity of the consolidated coding system. Of the five variants tested, version two (which included the need, expectation, affect, and outcome criteria for both HS and FF) performed best across the various indices of psychometric fitness. HS and FF scores derived from the second version were reasonably stable over time ($r = .344$, $r = .435$), and were found to correlate

strongly with their corresponding Heckhausen (1963b) scores ($r = .803$, $r = .703$), providing evidence of convergent validity. Furthermore, HS scores from version two predicted both speed ($p = .019$) and accuracy ($p = .044$) in the WCST. While version two FF scores did not significantly predict task speed in the BART, they did predict risk-taking after participants who responded extremely quickly were excluded from the analysis ($p = .052$).

Of the other system variants examined, version four performed reasonably well. The version four test-retest reliability statistics were superior to those of version two (HS, $r = .354$; FF, $r = .470$), while version four HS also predicted WCST outcomes more strongly (speed, $p = .016$; accuracy, $p = .028$). However, version four FF scores were only marginally predictive of BART risk-taking even after exclusion of rapid responders, meaning that version two offers a more balanced pattern of results. Furthermore, examination of the skewness and kurtosis of the motive scores derived from the five variants indicated the V4 FF was highly positively skewed. Version one exhibited unacceptably low test-retest reliability in the case of HS ($r = .286$, $p = .073$). While FF derived from versions three and five did not significantly predict risk-taking behaviour in the BART. As such, these system variants were discarded in favor of version two. To ensure that the coding rules associated with version two were applied judiciously, inter-rater reliability was calculated using the same procedure as in study 1B. 10% of the data were coded by a second rater who had previously exceeded 85% accuracy in their scoring of the practice stories. Inter-rater reliability was calculated on the basis of this subsample of data. The concordance rates for HS and FF were 89% and 88% respectively.

As in study 1B, the test-retest reliability statistics for both HS and FF fell short of

those obtained by Schultheiss and Pang (2007). Nonetheless, both the session one HS and FF scores derived from version two were significantly correlated with the scores obtained at time two, suggesting that the motive scores are stable over time. Furthermore, the FF scores obtained using version two of the consolidated system exhibited greater stability than the Heckhausen FF scores obtained in study 1. This finding can perhaps be attributed to the omission of several categories of questionable relevance to the motive to avoid failure, such as criticism and avoidant instrumental activity.

The FF predictive validity results of study 2C also differed from those obtained during study 1B. Whereas FF scores derived from Heckhausen (1963b) coding of the ambiguous picture set data predicted BART speed but not risk-taking, scores derived from the consolidated system coding of the same data displayed the opposite pattern, predicting risk-taking but not speed. While this pattern of results is confusing, it can be said that both results offer at least some support for the predictive validity of scores derived from the ambiguous variant of the new picture set, and partial support for the predictive validity of both the Heckhausen (1963b) and consolidated coding systems. Possible reasons for this divergent pattern of results will be considered in the general discussion.

GENERAL DISCUSSION

11. Summary of Findings

11.1 Study 1.

The first study was intended to develop the first independently validated PSE picture set for the assessment of hope of success and fear of failure, while at the same time investigating the role of a previously unstudied methodological consideration in picture cue selection: set ambiguity. In study 1A, two novel picture sets were developed and pre-tested, one that exhibited set ambiguity and one that did not. Individually, the images comprising the two picture set variants exhibited a good balance of pull and cue ambiguity (with respect to Heckhausen's HS and FF), while the resulting picture set variants both exhibited greater diversity in their depiction of gender, race, and achievement setting than two historically significant picture sets. These enhancements were expected to offer improvements in validity in study 1B.

Study 1B sought to compare the psychometric adequacy of these two picture sets. The test-retest reliability of the ambiguous variant was found to be superior, while both variants exhibited divergent validity. The two picture sets were also compared in terms of their ability to predict theoretically relevant behaviors, along with a third picture set comprising commonly used images from each of the major *nAch* assessment traditions. HS scores derived from the ambiguous variant predicted both increased speed and decreased accuracy in a time-dependent task, while FF scores were found to predict decreased speed in a time-independent task. Scores derived from the unambiguous and comparison picture sets did not significantly predict these same behaviors. Together, these results offer tentative support for the predictive validity of the ambiguous picture set

version, and suggest that the use of a diverse, contemporary picture set exhibiting set ambiguity offers substantial improvements in the reliability and validity of PSE assessment of HS and FF, while cautioning against the continuing use of older images.

11.2 Study 2.

Building on the development of an improved picture set in study 1, study 2 sought to further refine PSE assessment of the achievement motive by developing an enhanced content coding system for efficient and simultaneous measurement of HS, FF, and *n*Ach. In study 2A, cluster analysis was used to investigate the relationships between the McClelland et al. (1953), Heckhausen (1963b), Winter (1994), and Pang (2006) systems at both the motive and subcategory level. Results attested to the equivalence of *n*Ach and HS, while reinforcing the independence of FF from the approach oriented achievement motives. Furthermore, the results of study 2A identified instances of subcategory redundancy, motive contamination, and weak association with parent motives that would inform the construction of an improved, consolidated coding system in study 2B.

A first draft of the consolidated coding system was formalized and written in study 2B, using the template provided by the results of study 2A. Redundant categories were combined while unrelated categories were removed, and efficient scoring conventions were adopted so as to maximize the usability of the new system. In study 2C, the reliability and validity of the scores derived from this first draft system were compared with those derived from four other variants of the new system, each possessing fewer categories (and hence reduced complexity). Of all the different versions examined, version two—which included the need, expectation, affect, and outcome criteria for both

HS and FF—exhibited the best psychometric performance. HS and FF scores derived from version two possessed adequate test-retest reliability, and were found to exhibit both convergent and divergent validity. Critically, HS scores derived from this version were found to predict speed and accuracy in the same time-dependent task used in study 1, while FF scores from version two predicted decreased risk-taking, but not decreased speed, in the BART. These results offer some support for the predictive validity of the new coding system.

12. Picture Set Contributions

12.1. An overdue reappraisal.

The introduction to this thesis described an abbreviated history of the development and refinement of the PSE since its genesis in the mid twentieth century. During this time, much significant methodological advancement has been made, such as the development of new and improved content coding systems and the establishment of best-practices for administration (Pang, 2010b). However, while much effort has been expended on the refinement and redevelopment of content coding systems, appreciation of the critical role played by the pictures themselves has to some extent diminished. PSE research is now experiencing something of a renaissance, due to theoretical advancement and a greater choice in the number of available coding systems (Schultheiss & Brunstein, 2010), yet researchers persist in using images whose validity has been in steady decline for the best part of a century, and too few efforts have been made to address this pressing issue despite several calls to action (e.g., Schultheiss & Brunstein, 2001). As such, this research constitutes a timely shift in focus back to the pictures that are the very essence of the PSE, and it is hoped that other researchers in the field will be inspired to further examine the mechanics of picture cue selection while developing new image batteries for the assessment of other fundamental motives.

Furthermore, these results caution against the continuing, unreflective use of images from the McClelland and Heckhausen assessment traditions, given that a picture set comprising such images was found to exhibit low predictive validity, at least in terms of the WCST and BART, in study 1. While it cannot be denied that many researchers have gained important insight through continuing use of such images in PSE research over the

past decade (see Pang, 2010a for a review), this fact does not preclude the possibility that other phenomena have remained undetected given the declining validity of commonly used picture sets. In general, the use of newer, independently validated picture sets such as the ambiguous HS/FF picture set developed here, will afford PSE researchers more accurate measurement of implicit motives, and will thus facilitate knowledge generation in the field of implicit motivation. The specific contributions of the picture set developed herein will be discussed in turn in the following sections.

12.2. Improved relevance and decreased bias.

The results of study 1B vindicate the calls of researchers like Henry (1956) and Murstein (1972) for the continual updating of standard picture sets, and suggest that similar efforts should be made to update or replace picture sets typically used for the assessment of *nAff* or *nPow*. All images in the new picture set feature people dressed in a more contemporary style, engaging in activities that are personally relevant (or at least universally meaningful) to respondents in the present day, increasing the likelihood that participants will project their own unconscious states and desires onto the actors, rather than relying on schemas or tropes derived from popular culture. Recent personal experience of content coding stories derived from older pictures suggests the latter is an increasingly common trend. Several responses to Heckhausen's (1963b) *man at desk* picture (which was used as a comparison image in the present study), explicitly characterized the office worker as Don Draper from the popular television drama series *Mad Men*, incorporating themes from the show into the narrative. It seems probable that respondents reach for such associations when they feel no personal connection to the

individual or their situation, foregoing projection and undermining the validity of their PSE responses. The novel picture set avoids such pitfalls by including only individuals and situations that are relevant and universal in their meaning to modern day respondents.

A second improvement is in the diversity of the new picture set. Compared to older picture sets, which are almost universally male and Caucasian, the new picture set includes many depictions of women and individuals of other ethnicities. Improved representation of different ethnicities should increase the cross-cultural validity of PSE assessment of the achievement motive, and consequently may encourage researchers in non-Western cultures to adopt and use PSE methodology, addressing the much-lamented Western bias in psychology (e.g., Van De Vijver & Leung, 2000). The fact that the picture set was developed and validated in an East Asian sample adds further weight to claims of cross-cultural applicability. Furthermore, reducing ethnic bias in PSE picture cues may even confer improvements in validity when studying Western samples, since countries such as the United States have becoming increasingly diverse since the days of McClelland and colleagues (Massey, 2008), with non-Caucasian ethnic groups making up an ever larger percentage of the population.

12.3. Appropriateness for the assessment of HS and FF.

The new HS and FF picture set developed here finally presents researchers with the first picture set since Heckhausen's inaugural 1963 set specifically designed to assess HS and FF. The approach-avoidance distinction has become increasingly central to many influential models of the explicit achievement motive (e.g., Elliot, 1999; Elliot & Church, 1997), and research in this domain has flourished over the past couple of decades (Eder,

Elliot, & Harmon-Jones, 2013), yielding many valuable insights for the classroom and the workplace. Yet while research into explicit HS and FF has thrived, comparable investigation of implicit HS and FF has suffered from a loss of momentum. Few active researchers use the Heckhausen (1963b) system, and those who do persist with the use of older picture sets whose validity the present research has called into question (e.g., Gruber & Kreuzpointner, 2013). By developing and validating a picture set specifically for the assessment of HS and FF, one that demonstrates adequate pull for both motives and gives rise to scores with sufficient variance for the instrument to possess predictive validity, it is hoped that the present research will encourage implicit motive researchers to assess and incorporate the HS/FF distinction into their investigations of the achievement motive. In combination with the newly developed content coding system (the unique contributions of which will be discussed in more detail in the following sections), the new HS/FF picture offers a valuable new option for researchers in the field.

12.4. Independent picture set validation.

In addition, the methods adopted in the present research address the lack of independent validation that afflicts many picture sets. The lack of attention paid to the picture set component of the PSE is reflected in many discussions of PSE validity, which typically refer to the validity of the coding system when, in fact, they should more accurately refer to the validity of the combined picture set and coding system instrument. Typically, picture sets are viewed as accessories to the coding system rather than an integral part with a distinct contribution to the instruments overall psychometric performance, meaning that few attempts are ever made to validate the picture sets

specifically (although see Pang, Cheng, & Lim, 2008). Consistent with one of the primary aims of this research—to reestablish the picture set as an important topic of methodological enquiry in PSE research—study 1 was structured in such a way that the reliability and validity of the new picture set was assessed independently, so that any observed improvements could be attributed to one component alone. Consequently, the new picture set represents the first and only image battery developed specifically for HS/FF assessment that has been validated independently from an associated coding system.

12.5. The importance of set ambiguity.

Another important contribution of the present research is the identification of a previously unknown methodological consideration in picture set construction—that of set ambiguity—and the concomitant development of a picture set that exhibits this essential property. Although a number of important questions still need to be addressed (discussed in more detail in section 15), these findings reaffirm the importance of taking a holistic approach to picture cue selection, by attending not only to the properties of individual cues but also to the composition of the overall picture set. Critically, they suggest that all PSE picture sets, whether constructed for the purpose of single-motive or multi-motive assessment, should be varied in terms of the individual picture cues' motivational focus. While implicit motive researchers have long espoused the benefits of cue variability when studying multiple motives (e.g., Pang, 2010a), the prevailing wisdom has been that single-motive assessment is best conducted using uniform picture sets consisting entirely of images that pull preferentially for the motive under investigation (e.g., Smith et al., 1992).

The present research suggests that this is not the case. Inclusion of two cues that pull preferentially for an alternative motive alongside a majority of cues that pulled for HS and FF afforded substantial improvement in the reliability and validity of the resulting scores. Given the improvements in reliability and validity of the HS and FF scores of the ambiguous picture set relative to the unambiguous picture set, I therefore recommend that researchers construct ambiguous picture sets when conducting PSE research.

13. Coding System Contributions

13.1. A pathway to integration.

The introduction to this thesis highlighted the issue of fragmentation in field of implicit achievement motivation, and suggested that this phenomenon has hampered progress within the field and hindered integration with the mainstream of motivational psychology. This fragmentation is driven by the parallel use of different content coding systems in semi-distinct literatures, with few attempts to empirically establish the equivalence of the measures, target constructs or resulting findings. The impracticalities associated with mastering and applying multiple coding methodologies is likely to be a major contributing factor. Implicit motive researchers are rarely inclined to master multiple coding systems, and prefer instead to stick to their tried and tested methods, usually the multipurpose Winter (1994) *Manual for Scoring Motive Imagery in Running Text*.

To compound this issue, each of the most widely used content coding systems for the PSE assessment of the implicit achievement motive have significant drawbacks and

weaknesses, and since the development and refinement of the Winter system in the early nineties, only Pang (2006) has sought to push the field forward by developing a new and improved PSE coding system for implicit achievement motive assessment. The present research builds on the foundations of Pang (2006), providing a simplified system for the simultaneous assessment of *nAch*/HS and FF. In adopting this system, researchers will no longer need to choose whether to measure the classical, unitary *nAch* construct or to adopt a less well-known and potentially more time-consuming measure that incorporates the HS/FF distinction.

An important theoretical contribution of the present research is its confirmation of the equivalence of *nAch* and HS. Rather than clustering together with HS-related variables in a superordinate *nAch* cluster, categories and motive constructs associated with FF clustered together in study 2A, suggesting that the motive to avoid failure is distinct and independent. On the other hand, categories and motives associated with HS and *nAch* clustered together, suggesting that the motive to approach success and the need for achievement are one and the same. While this finding raises issues of motive contamination in both the McClelland et al. (1953) and Winter (1994) systems (which will be addressed in section 13.4), the method of construction of the consolidated coding system ensures compatibility between the new system and the systems from which it was derived. It is hoped that these improvements, both methodological and theoretical, will allow for greater cross-talk between the *nAch* and HS/FF research traditions, while encouraging more researchers to incorporate the approach-avoidance distinction into their work on the achievement motive.

13.2. Incorporation of the distinction between HS and FF.

To a large extent, the major contributions of the new coding system mirror those of the new picture set. As described in section 12.3, implicit motive researchers have been surprisingly reluctant to adopt the Heckhausen (1963b) and Pang (2006) systems despite the more nuanced understanding that the separate measurement of HS and FF affords. This state of affairs stands in stark contrast to the explicit motive literature, where the approach-avoidance distinction is central to many prominent theories of explicit achievement motivation (Elliot & Harackiewicz, 1994, Elliot & Church, 1997, Covington & Omelich, 1991, Covington & Roberts, 1994). This prominence has given rise to a vibrant and active literature on the approach and avoidance in explicit achievement motivation, with several important new findings published in the last year alone (e.g., Bélanger et al., 2013, Oertig et al., 2013).

It is hoped that the present research will help redress this situation. By constructing a need achievement PSE content coding system that incorporates the approach-avoidance distinction while offering the simplicity of administration associated with the Winter (1994) system, this research offers an appealing alternative for implicit motive researchers studying the achievement motive. While the Winter system's ability to simultaneously code for the big three motives still represents an advantage (see section 15 for further discussion), the new system represents a more refined and more comprehensive measure of the achievement motive.

13.3. Simplification of administration.

The simplification of the present system takes many forms. Firstly, the need to

decide whether the story features any achievement-related imagery before scoring commences, as is required by the McClelland et al. (1953) system has been discarded, following the examples set by Heckhausen (1963b), Winter (1994), and Pang (2006). Abandonment of this convention means the scorer saves time by moving straight into line-by-line coding without having to ponder over macro-level questions regarding the overall achievement-relatedness of the narrative. Secondly, the new system follows Winter's example by eschewing subcategories in favor of criteria for scoring the parent motive. While researchers employing the McClelland et al. (1953) or Heckhausen (1963b) systems are frequently forced to consult the literature when attempting to decide whether a given phrase corresponds better to one category or another, adopting a criteria system removes the need for these time-consuming consultations. Thirdly, the number of criteria associated with each motive has been kept to a minimum. While the first draft of the consolidated system constructed in study 2B featured six HS criteria and five FF criteria, the finalized version derived from study 2C featured only eight criteria, four for each motive. Finally, small improvements such as the inclusion of detailed explanations regarding scoring decisions in the practice stories should assist in learning and mastering the system, further reducing the barriers to adoption.

13.4. Avoidance of motive contamination.

Other than the incorporation of the approach-avoidance distinction, one of the biggest advantages the new system has over its predecessors is the avoidance of motive contamination. As outlined in the introduction, motive contamination is rife in the McClelland et al. (1953) system, given that AI is often scored for achievement pursued for

the sake of reputational enhancement (*nPow*) and *Nup* is scored for benevolent assistance (*nAff*). Most importantly, the McClelland et al. (1953) system, which is essentially a measure of the approach-related need for achievement (i.e., *HS*), is contaminated by the inclusion of categories that are more closely related to *FF*, such as *Ga-*, *G-*, *Bw*, and *Bp*.

While later systems did much to address these deficiencies, several suffer from issues of motive contamination in their own right. The Winter (1994) system obscures the *HS/FF* distinction, collapsing McClelland et al.'s (1953) subcategories into five criteria, four of which are approach-related while the other relates to *FF*. As such, the Winter (1994) measure of *nAch/HS* is contaminated with categories pertaining to the need to avoid failure. In the case of the Pang (2006) system, the results of study 2A suggest that the category *QCO* may be more indicative of power motivation. The methods adopted in study 2, where categories from the various systems were clustered along with *nPow* and *nAff*, ensured that any instances of affiliation or power contamination could be identified, while the two-cluster *HS/FF* solution reaffirms the dangers of conflating the approach-avoidance distinction.

14. Wider Contributions

The new consolidated coding system and associated picture set have many practical applications, particularly in domains of employment and education. As outlined in the introduction, it is widely-acknowledged that implicit and explicit motives are fundamentally distinct, predicting different classes of behaviors in response to different task-related or situational cues (McClelland, Koestner, & Weinberger, 1989), yet few recruiters or human resource professionals make use of projective measures like the *PSE* during selection and assessment (the notable exception being the Hay Group, who

acquired the David McClelland founded consultancy firm McBer in 1984). Most recruitment consultancies and HR departments prefer to focus on explicit measures of achievement orientation, such as achievement subscale of the Jackson (1974) PRF or the achievement motive inventory (AMI; Schuler, Thornton, Frintrup, & Mueller-Hanson, 2004), given their administrative convenience. As a consequence, organizations may be failing to identify those individuals who are most motivated to work because of the intrinsic challenge associated with their job, instead targeting those who are motivated to perform by external incentives. The development of a simplified, enhanced content coding system and associated picture set for the simultaneous assessment of *nAch*/HS and FF offers employers and recruiters a powerful new tool for identifying talent.

The ability of the new system to code for FF, in conjunction with the development of the only contemporary picture set specifically designed for the assessment of this motive, should be of particular value for recruitment and human resource management. While the notion that fear of failure is entirely debilitating with respect to achievement motivation has to a large extent been dispelled (Pang, 2010a), there can be little doubt that certain jobs and working environments are less suitable for those motivated primarily by FF. On the other hand, the more methodical, risk-averse pattern of responding observed in the present study might render FF-motivated individuals more suitable for roles requiring caution or vigilance, such as airline baggage checker or insurance underwriter. The new content coding system and picture set offer employers and recruiters an enhanced tool for measurement of both *nAch*/HS and FF, and should assist in their efforts to match individuals to jobs on the basis of their motivational profile.

The new coding system and picture set also have the potential to contribute in the

classroom. While the PSE is rarely used to profile students in schools, it possesses many qualities that render it highly suitable for use with younger individuals. Creative writing can be intrinsically rewarding, and many students might relish the opportunity to engage in an ungraded exercise where they are free to express themselves without worrying about their performance relative to others. The PSE does not look or feel like a test in any conventional sense, and it is likely that youths would respond in a more naturalistic (and hence more valid) way to a more subtle instrument such as this. Furthermore, PSE administration can easily be framed as a creative writing exercise to minimize the potential anxiety associated with tests or questionnaires. Given these appealing properties, the new coding system and picture set, with all the previously described improvements in reliability, validity and efficiency, should be able to identify FF-motivated students in a subtle, non-threatening way, allowing them access to the additional assistance they may require.

This recommendation for the use of the new picture set and consolidated coding system in employment and education settings comes with a necessary caveat. Diagnostic use of the PSE or any other measure of personality in an applied setting first requires the establishment of population norms for different genders, ages, and cultural or ethnic groups, without which meaningful comparisons between employees or students cannot be made. The lack of adequate norms for projective measures used in clinical settings has been criticized by several researchers (e.g., Lilienfeld, Wood, & Garb, 2000), and this issue must be addressed if the combined picture set and coding system measure is to be used in this manner.

15. Limitations and Future Directions

Several limitations of the present research should be noted. In both studies one and two, one of the hypotheses with regard to the enhanced predictive validity of the new instrument was not borne out by the data. In study 1B, the hypothesized relationship between FF derived from the ambiguous picture set variant and risk-taking in the BART was not observed, while in study 2C, the significant relationship between FF and speed in the BART, that was initially observed in study 1B, was not replicated. These results are contradictory, and further experimentation will be required to identify whether suboptimal task choice or inadequate statistical power explain these null results. In study 2C, which should offer the greatest predictive validity since the new coding system and picture set are being used in tandem, exclusion of rapid responders did mean that the p value associated with FF derived from coding system version two moved closer to significance (from $p = .21$ to $p = .17$, with a positive value for β). Sample size was heavily restricted in study 2C ($N = 41$) since it would have been inappropriate to reanalyze data from the unambiguous picture set variant, as this picture set had already been shown to be inadequate in terms of both reliability and validity, and this already small number was further reduced when excluding the rapidly responding participants. This small sample size represents a significant limitation of the study providing evidence for the validity of the consolidated coding system. Consequently, a replication of study 2C with a much larger sample size is necessary to resolve the issues surrounding the predictive validity of FF scores derived from both the new picture set and the consolidated coding system. Hypothetically, if the null or inconsistent results were found to persist even after a substantial increase in statistical power, new validation tasks would have to be found to

more rigorously test the validity of FF scores derived from the new coding system and picture set. The relative paucity of research describing the behavioral correlates of fear of failure makes identifying suitable validation criteria quite challenging, and stands in stark contrast to the wealth of evidence describing the behavior of individuals high in *nAch*/HS. Nonetheless, the validation results described in this thesis are encouraging, offering tentative new insight into the behavior of FF motivated individuals. The results of study 1B suggest that FF-motivated individuals work more slowly on challenging tasks, but only when freed from external time pressure, while the results of study 2C suggest that the achievement-related risk aversion that characterizes FF-motivated individuals also extends to other domains, since FF scores were found to associated with a generalized measure of risk-taking. These results further characterize the motive to avoid failure, while at the same time attesting to the predictive validity of the assessment system developed to measure this heretofore poorly characterized motive.

Another limitation of the present research concerns the number of repetitions of PSE administration. By asking participants to write stories in response to PSE images depicting achievement-related activity on three separate occasions, the present study risked boring participants or increasing the likelihood of them discerning the purpose of the experiment. The author's personal experience suggests that some participants become frustrated when completing the PSE more than once, while the argument made in this thesis regarding set ambiguity—that responding to too many achievement images may risk arousing explicit achievement motivation—also applies to repetition of achievement images over multiple sessions. While the administration of three PSEs during the present study was necessary to test hypotheses regarding test-retest reliability and engage in

validity comparisons, it should be noted that such a design is not without its drawbacks.

Beyond the WCST and BART, further work is required to comprehensively validate both *n*Ach/HS and FF scores derived from the new coding system and picture set. Of the other coding systems available, the McClelland et al. (1953), Heckhausen (1963b), and Winter (1994) systems have all been widely used in achievement motive research, and so have been extensively validated with respect to multiple behavioral criteria. While the consolidated coding system offers significant improvements over these systems, the fact that all data supporting the validity of the new coding system are derived from a limited number of studies using a restricted array of behavioral correlates represents a further limitation of the present research. All of the studies supporting the validity of the new picture set and coding system are correlational in nature. Further research, using both empirical, laboratory-based designs and quasi-experimental longitudinal designs examining real-life outcomes, will be required to comprehensively validate the new picture set and consolidated coding system.

Furthermore, one of the primary aims of study 1 was to develop a more diverse picture set featuring a range of ethnic groups, with the hope that these improvements would lead to heightened validity of motive assessment borne of enhanced relevance and stimulus pull. In essence, it was hoped that the new picture set would exhibit cross-culturally validity, although it is acknowledged that this hypothesis was not directly tested during the present research. Each of the studies in this thesis were conducted in similar samples derived from the same Asian university, and so the validity of the system with respect to motive assessment in other developed nations and cultures remains an open, empirical question. Future research should seek to establish the reliability and validity of

the picture set in culturally distinct samples. It is anticipated that the pretesting and selection methods adopted in study 1A will ensure that new picture set performs well in this regard, although it is likely that there will be limitations to this cross-cultural validity. For example, it is unlikely that the new picture set will be equally effective in developing nations, where the achievement scenarios and styles of dress depicted in the new picture set are neither relevant nor universal. This constitutes another limitation of the present research.

A further limitation of the present research concerns the suitability of the new picture set to motive assessment among different age groups. Existing guidelines for the development of new picture sets advise that selected images should “not depict recognizably extreme examples (e.g., characters in the pictures are neither very young nor very old)” (Pang, 2010b, p. 133). While I believe the new picture set fulfills this criterion, the fact that it was developed and validated in an undergraduate sample means that it likely be most effective at measuring HS and FF in similar samples. Nonetheless, the absence of recognizably extreme examples, such as young children and the elderly, from the picture set suggests that the new picture set should exhibit reasonable generalizability across a range of age groups.

With respect to the importance of set ambiguity in picture set construction, further empirical work should be conducted in order to replicate and extend these preliminary findings. Several areas warrant particular attention. Firstly, researchers should replicate these findings in context of other content coding systems with varying motivational foci, such as McAdams (1980), Winter (1973), and Winter (1994). While it seems likely that the principle of set ambiguity should apply equally to these alternate systems and motives,

empirical confirmation of these suspicions is required. Secondly, research is needed to identify the optimal ratio of primary to alternate motive pull images for single-motive assessment batteries. While the present research demonstrates the advantage of including two alternate pull cues in an eight picture set, the possibility remains that including more alternate pull cues—that is, increasing set ambiguity still further—offers even greater improvements in reliability and validity. A study in which set ambiguity is varied incrementally, conducted in a manner similar to Schultheiss and Pang's (2007) investigation of picture set size, would shed light on this issue. Furthermore, it is currently unknown whether inclusion of alternate motive pull images is a strict requirement for ensuring reliability and validity in single-motive PSE assessment, since including a number of neutral pictures with low pull for all motives may also suffice. A picture set comprising both target motive pull images and neutral, low pull images would also exhibit set ambiguity, and future research should attempt to establish whether neutral images can serve the same purpose as the alternate motive pull images used here. Finally, further work is needed in order to explain exactly why set ambiguity improves the psychometric properties of the PSE. Although I tentatively suggested in my introduction that unambiguous picture sets might exert undesirable effects due to demand characteristics, no evidence has been gathered that directly supports this contention. No participants indicated their awareness of the experimental hypothesis in their responses to funneled debriefing questions in the present study, and unambiguous HS and FF were not found to be significantly correlated with explicit HS and FF respectively, although correlations between average HS and FF and their explicit counterparts were more positive in the unambiguous condition. Future research should explore this and other possible

explanations more fully.

One obvious and very important way that this work can be extended is through similar methodological refinement of the systems and pictures commonly used to assess the two remaining members of the big three motivational trinity: *nPow* and *nAff*. This research is vital for several reasons. Firstly, the pictures commonly used for PSE assessment of these motives are as just as dated as those used to assess *nAch*, and so are almost certain to suffer from the same issues of declining validity. Secondly, the approach- and avoidance-related manifestations of *nAff* and *nPow* have been studied far less frequently than HS and FF (particularly in the case of *nPow*), a situation that reflects the status of *nAch* as the most extensively studied of all higher-order motives, and so no PSE coding systems or stimulus images have been developed specifically to test for hope of affiliation (HA; the approach manifestation of *nAff*) fear of rejection (FR; the avoidant manifestation of *nAff*), hope for power (HP; the approach manifestation of *nPow*), and fear of power (FP; the avoidant manifestation of *nPow*)¹. Thirdly, the Winter (1994) multimotive system allows for simultaneous coding of *nAch*, *nPow*, and *nAff*, and will likely to continue to be the coding system of choice for many implicit motive researchers unless an alternative system is developed that offers the same level of convenience by simultaneously coding for the *nAch*, *nAff*, and *nPow*. Without such a system, it is likely

¹ It is acknowledged that the Boyatzis (1972; 1973) system does code separately for HA and FR, yet as a non-pictorial measure that requires respondents to complete word stems, it does not qualify as a true PSE measure of the approach- and avoidance-related manifestations of *nAff*.

that the fundamental approach-avoidance distinction in motivation will continue to languish on the sidelines of implicit motive research, rather than deservedly taking center stage as it has done in the explicit achievement goal literature.

The new picture sets should be developed using methods similar to those employed in study 1, although independent validation of the new image batteries will not be possible given the absence of previously validated content coding systems for the assessment of HA/FR and HP/FP. Firstly, a comprehensive effort should be made to pretest and select contemporary images that can then be used with each other and with images from the new HS/FF picture set to construct ambiguous picture sets for the assessment of these motives. Construction of the content coding systems will represent more of a challenge. In the absence of preexisting systems incorporating the approach-avoidance distinction¹, categories from the original coding systems for *nAff* (Atkinson, Heynes, & Veroff, 1954) and *nPow* (Veroff, 1957), along with those derived from informative successor measures (Boyatzis, 1972; 1973; Uleman, 1972; Winter, 1973) should firstly be respecified into separate positive (approach) and negative (avoidance) categories following the examples of Clark, Teevan, and Ricciuti (1956), Moulton, (1956), Heckhausen (1963b). Subsequently, a PSE response corpus should then be coded with adequate reliability using each of these respecified systems and the Winter (1994) system, with the resulting data being subjected to the same kind of cluster analysis employed in study 2A. Assuming that a stable four cluster solution is obtained, the results of the cluster analysis can then be used to inform the construction of a consolidated coding system for the assessment of HA, FR, HP, and FP, which can then be combined with the new consolidated HS/FF system to form the first PSE content coding system for the

simultaneous assessment of the approach and avoidant manifestations of the big three motives.

At this stage, it would be remiss not to refer back to two other measures that go some way to fulfilling this objective: the MMG and the OMT. Building on work in the TAT/PSE tradition, the MMG requires participants to judge a series of 12 motivationally-relevant line drawings, such as a couple playing badminton or a group of work colleagues holding a round-table discussion. However, unlike the PSE, in which participants are required to spontaneously generate narratives without further advice or instruction, the MMG provides participants with a series of 14 motive-related statements corresponding to one of the six motives (e.g., “thinking about lacking abilities at this task” (FF) and “trying to influence other people” (HP), and asks them to judge how well each of these statements capture the content of the various images. The MMG was originally developed to combine of both PSE and self-report assessment methods, and it is commonplace to refer to the MMG as a *semiprojective* measure given its reliance on the endorsement of declarative statements. As a hybrid measure, it is perhaps unsurprising that many researchers don't consider the MMG to be a true measure of implicit motivation, but rather as one that assesses a combination of implicit and explicit motives. Brunstein and Schmitt (2003) found PSE and MMG measures of HS and FF to be uncorrelated, while Kehr (2004) observed weak but significant correlations between MMG scores and explicit measure of the big three motives. Consequently, while MMG scores have been found to predict task enjoyment (Puca & Schmalt, 1999), intrinsic motivation (Sokolowski & Kehr, 1999), and other variables related to implicit motivation, it is doubtful that the MMG measures implicit motives in isolation, underlining the need for a truly projective, PSE-

based measure of the approach- and avoidance-related manifestations of *nAch*, *nAff*, and *nPow*.

The OMT is perhaps closer to a true measure of implicit motivation. Participants are presented with a series of 15 images and asked to write down their “spontaneous associations” (Kuhl, 2008, p. 302) relating to the following questions: “What is important for the person in this situation and what is he or she doing? How does the person feel? Why does the person feel this way?”. The time taken to administer each image in the OMT is reduced given that participants are not required to form coherent narratives, yet the test retains the projective, non-declarative qualities of the TAT/PSE since the participants are not endorsing pre-written statements. Furthermore, the OMT coding system allows for simultaneous coding of the three big three motives, while further subdividing each motive into five mechanisms corresponding to four forms of approach motivation and one of avoidance motivation. Comparison of scores derived from the OMT with traditional implicit and explicit measure tentatively suggest it is an implicit measure. Scores from the OMT correlate with PSE-derived measures, but only when the arousal conditions specific to the motive under investigation are induced (Scheffer, 2000; Scheffer et al., 2003), while OMT scores have been found not to correlate with those derived from explicit measures (Scheffer, 2003). However, recent data from Schüller, Wegner, and Brandstätter (2013) revealed very low correlations between scores derived from the PSE and OMT, once again casting doubt on whether these tests truly measure the same constructs. Given the rich history of development and validation possessed by the PSE, it remains the most effective tool for the assessment of implicit motivation, underlining the importance of the work described in this thesis and the suggestions for

future research described in the present section.

16. Conclusions

The present research offers a complete and thorough refinement of the Picture Story Exercise (PSE) method of assessment for the implicit need for achievement (*nAch*). In study 1, the first contemporary picture set specifically designed for the assessment of the approach- and avoidance-related manifestations of *nAch*—termed hope of success (HS) and fear of failure (FF) respectively—was developed and subsequently validated in conjunction with the Heckhausen (1963b) content coding system for HS and FF. HS and FF scores derived from the ambiguous variant of the new picture set were found to be reliable, and exhibited better predictive validity than scores derived from either the unambiguous variant or a set of commonly used images. The new picture set offered improvements in terms of enhanced relevance and decreased bias, validation independent of an associated coding system, and identified the importance of a previously undocumented methodological consideration, termed set ambiguity, when constructing PSE picture sets. More generally, study 1 attempted to redress an imbalance in PSE research, stressing the importance of considering not only the content coding systems but also the picture sets.

A consolidated content coding system for the simultaneous assessment of *nAch*, HS, and FF was developed and validated during study 2. Cluster analysis of content coding data derived from the McClelland et al. (1953), Heckhausen (1963b), Winter (1994), and Pang (2006) systems identified instances of redundancy, motive contamination, and strong versus weak parent motive associations amongst the various

subcategories, with these results subsequently guiding the development of a consolidated system that retained the best features of the constituent systems, including efficiency of administration, lack of motive contamination, and incorporation of the HS/FF distinction. Scores derived from the new system were found to be stable over time, and possessed convergent, divergent, and predictive validity. Study 2 offered empirical confirmation of the equivalence of *n*Ach and HS using a novel method, and reaffirmed the validity and theoretical importance of the approach-avoidance distinction in implicit achievement motivation research.

To conclude, it is hoped that the present findings and methodological advancements will prompt more *n*Ach researchers to frame research questions in terms of the HS/FF distinction, so that the contrast between approach and avoidance motivation takes its rightful place at the heart of implicit motive research. It is also hoped that the simplicity and effectiveness of the new system will encourage those outside the field to adopt PSE methodology in their study of achievement motivation. The TAT/PSE has a long and distinguished history, and it is only through continual refinement and reappraisal that this unique instrument can retain its power to uncover the secrets of the unconscious mind.

REFERENCES

- Ach, N. (1910). *Über den Willensakt und das Temperament*. Leipzig: Quelle & Meyer.
- Anderson, J. W. (1999). Henry A. Murray and the creation of the Thematic Apperception Test. In L. Geiser & M. I. Stein (Eds.), *Evocative Images: The Thematic Apperception Test and the Art of Projection* (pp. 23-38). Washington, DC: American Psychological Association.
- Aron, B.R. (1949). *A manual for analysis of the Thematic Apperception Test: A method and technique for personality research*. Berkeley, CA: WE Berg.
- Aron, B.R. (1950). The Thematic Apperception Test in the study of prejudiced and unprejudiced individuals. In T. W. Adorno, E. Frenkel-Brunswick, D. J. Levinson & R. N. Sanford (Eds.), *The authoritarian personality* (pp. 489-544). New York, NY: Harper & Row.
- Atkinson, J.W. (1947). *The functional relationship between an individual's needs and his values*. Unpublished honors thesis, Wesleyan University. Middletown, CT.
- Atkinson, J.W. (1950). *Studies in projective measurement of achievement motivation*. Unpublished doctoral dissertation, University of Michigan. Ann Arbor, MI.
- Atkinson, J.W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, 64(6), 359-372. doi: 10.1037/h0043445
- Atkinson, J. W. (1958). *Motives in fantasy, action, and society: A method of assessment and study*. Princeton, NJ: Van Nostrand.
- Atkinson, J. W. (1992). Motivational determinants of thematic apperception. In C. P. Smith (Ed.), *Motivation and personality: Handbook of thematic content analysis* (pp. 21-48). New York, NY: Cambridge University Press.

- Atkinson, J. W., & Birch, D. (1970). *The dynamics of action*. Oxford, UK: John Wiley.
- Atkinson, J. W., Heyns, R. W., & Veroff, J. (1954). The effect of experimental arousal of the affiliation motive on thematic apperception. *The Journal of Abnormal and Social Psychology*, 49(3), 405-410. doi: 10.1037/h0053499
- Atkinson, J. W., & Litwin, G. H. (1960). Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. *The Journal of Abnormal and Social Psychology*, 60(1), 52-63. doi: 10.1037/h0041119
- Baard, P. P., Deci, E. L., & Ryan, R. M. (2004). Intrinsic Need Satisfaction: A Motivational Basis of Performance and Well-Being in Two Work Settings. *Journal of Applied Social Psychology*, 34(10), 2045-2068. doi: 10.1111/j.1559-1816.2004.tb02690.x
- Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50(1-3), 7-15. doi: 10.1016/0010-0277(94)90018-3
- Bélanger, J. J., Lafrenière, M.-A. K., Vallerand, R. J., & Kruglanski, A. W. (2013). When passion makes the heart grow colder: The role of passion in alternative goal suppression. *Journal of Personality and Social Psychology*, 104(1), 126-147. doi: 10.1037/a0029679
- Berglas, S., & Jones, E. E. (1978). Control of attributions about the self through self-handicapping strategies: The appeal of alcohol and the role of underachievement. *Personality and Social Psychology Bulletin*, 4(2), 200-206. doi: 10.1177/014616727800400205

- Berlew, D. E. (1956). *The achievement motive and the growth of Greek civilization*. Unpublished honors thesis, Wesleyan University. Middletown, CT.
- Birney, R. C. (1958). Thematic content and the cue characteristics of pictures. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society* (pp. 630-643). Princeton, NJ: Van Nostrand.
- Bouffard, T., Boisvert, J., Vezeau, C., & Larouche, C. (1995). The impact of goal orientation on self-regulation and performance among college students. *British Journal of Educational Psychology*, 65(3), 317-329. doi: 10.1111/j.2044-8279.1995.tb01152.x
- Boyatzis, R. E. (1973). Affiliation motivation. In D. C. McClelland & R. S. Steele (Eds.), *Human motivation: A book of readings* (pp. 252-276). Morristown, NJ: General Learning.
- Boyatzis, R. E. (1973). *A two-factor theory of affiliation motivation*. Unpublished doctoral dissertation, Harvard University. Cambridge, MA.
- Brickenkamp, R., & Zillmer, E. (1998). The d2 test of attention. Göttingen: Hogrefe & Huber.
- Brunstein, J. C., & Heckhausen, H. (2008). Achievement motivation. In J. Heckhausen & H. Heckhausen (Eds.), *Motivation and action* (pp. 137-183). New York, NY: Cambridge University Press.
- Brunstein, J. C., & Maier, G. W. (2005). Implicit and Self-Attributed Motives to Achieve: Two Separate but Interacting Needs. *Journal of Personality and Social Psychology*, 89(2), 205-222. doi: 10.1037/0022-3514.89.2.205

- Brunstein, J. C., & Schmitt, C. H. (2003). *Prüfung der konvergenten, diskriminanten und prädiktiven Validität von Leistungsmotiv-IATs, -TATs und -Fragebögen [Testing the convergent, discriminant, and predictive validity of the achievement motive IAT, TAT, and questionnaires]*. Unpublished work, University of Potsdam. Potsdam, Germany.
- Brunstein, J. C., & Schmitt, C. H. (2004). Assessing individual differences in achievement motivation with the Implicit Association Test. *Journal of Research in Personality*, 38(6), 536-555. doi: 10.1016/j.jrp.2004.01.003
- Brunstein, J. C., Schultheiss, O. C., & Grässman, R. (1998). Personal goals and emotional well-being: The moderating role of motive dispositions. *Journal of Personality and Social Psychology*, 75(2), 494-508. doi: 10.1037/0022-3514.75.2.494
- Carver, C. S. (2006). Approach, avoidance, and the self-regulation of affect and action. *Motivation and Emotion*, 30(2), 105-110. doi: 10.1007/s11031-006-9044-7
- Chusmir, L. H., & Azevedo, A. (1992). Motivation needs of sampled Fortune-500 CEOs: Relations to organization outcomes. *Perceptual and Motor Skills*, 75(2), 595-612. doi: 10.2466/pms.75.5.595-612
- Clark, R., & Sensibar, M. (1955). The relationship between symbolic and manifest projections of sexuality with some incidental correlates. *Journal of Abnormal and Social Psychology*, 50(3), 327-334. doi: 10.1037/h0047775
- Clark, R. A., Teevan, R., & Ricciuti, H. N. (1956). Hope of success and fear of failure as aspects of need for achievement. *The Journal of Abnormal and Social Psychology*, 53(2), 182-186. doi: 10.1037/h0046496

- Clatworthy, J., Buick, D., Hankins, M., Weinman, J., & Horne, R. (2005). The use and reporting of cluster analysis in health psychology: A review. *British Journal of Health Psychology*, 10(3), 329-358. doi: 10.1348/135910705X25697
- Collins, C.J., Hanges, P.J., & Locke, E.A. (2004). The relationship of achievement motivation to entrepreneurial behavior: A meta-analysis. *Human Performance*, 17(1), 95-117. doi: 10.1207/S15327043HUP1701_5
- Combs, A.W. (1946). A method of analysis for the Thematic Apperception Test and autobiography. *Journal of Clinical Psychology*, 2(2), 167-174. doi: 10.1002/1097-4679(194604)2:2<167::AID-JCLP2270020214>3.0.CO;2-J
- Covington, M. V., & Omelich, C. L. (1985). Ability and effort valuation among failure-avoiding and failure-accepting students. *Journal of Educational Psychology*, 77(4), 446-459. doi: 10.1037/0022-0663.77.4.446
- Covington, M. V., & Omelich, C. L. (1991). Need achievement revisited: Verification of Atkinson's original 2 x 2 model. In C. D. Spielberger, I. G. Sarason, Z. Kulcsár & G. L. van Heck (Eds.), *Stress and emotion: Anxiety, anger, and curiosity* (Vol. 14, pp. 85-105). Washington, DC: Hemisphere.
- Covington, M. V., & Roberts, B. W. (1994). Self-worth and college achievement: Motivational and personality correlates. In P. R. Pintrich, D. R. Brown & C. E. Weinstein (Eds.), *Students motivation, cognition, and learning* (pp. 157-188). Hillsdale, NJ: Erlbaum.
- de Charms, R., & Carpenter, V. (1968). Measuring motivation in culturally disadvantaged children. In H. Klausmeirer & G. O'Hearn (Eds.), *Research and development toward the improvement of education*. Madison, WI: Educational Research

Services.

- de Charms, R., & Dave, P. N. (1965). Hope of success, fear of failure, subjective probability, and risk-taking behavior. *Journal of Personality and Social Psychology, 1*(6), 558-568. doi: 10.1037/h0022021
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum.
- Donley, R. E. (1968). *Psychological motives and the American Presidency*. Unpublished honors thesis, Wesleyan University. Middletown, CT.
- Donley, R. E., & Winter, D. G. (1970). Measuring the motives of public officials at a distance: An exploratory study of American presidents. *Behavioral Science, 15*(3), 227-236.
- Dugard, P., Todman, J. B., & Staines, H. (2009). *Approaching multivariate analysis: a practical introduction*. London, UK: Routledge.
- Dweck, C. S., & Elliot, A. J. (1983). Achievement Motivation. In E. Heatherington (Ed.), *Handbook of Child Psychology* (pp. 52-72). New York, NY: Wiley.
- Eder, A. B., Elliot, A. J., & Harmon-Jones, E. (2013). Approach and avoidance motivation: Issues and advances. *Emotion Review, 5*(3), 227-229. doi: 10.1177/1754073913477990
- Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational psychologist, 34*(3), 169-189. doi: 10.1207/s15326985ep3403_3
- Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology, 72*(1), 218-232. doi: 10.1037/0022-3514.72.1.218

- Elliot, A. J., & Covington, M. V. (2001). Approach and avoidance motivation. *Educational Psychology Review*, 13(2), 73-92. doi: 10.1023/A:1009009018235
- Elliot, A. J., & Dweck, C. S. (2005). Competence and Motivation: Competence as the Core of Achievement Motivation. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation*. (pp. 3-12). New York, NY US: Guilford Publications.
- Elliot, A. J., & Harackiewicz, J. M. (1994). Goal setting, achievement orientation, and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology*, 66(5), 968-980. doi: 10.1037/0022-3514.66.5.968
- Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, 80(3), 501-519. doi: 10.1037/0022-3514.80.3.501
- Elliot, A. J., McGregor, H. A., & Thrash, T. M. (2002). The need for competence. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research*. (pp. 361-387). Rochester, NY US: University of Rochester Press.
- Elliot, A. J., & Sheldon, K. M. (1997). Avoidance achievement motivation: a personal goals analysis. *Journal of Personality and Social Psychology*, 73(1), 171-185. doi: 10.1037/0022-3514.73.1.171
- Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion*, 32(3), 158-172. doi: 10.1007/s11031-008-9102-

- Entwisle, D. R. (1972). To Dispel Fantasies About Fantasy-Based Measures of Achievement Motivation. *Psychological Bulletin*, 77(6), 377-391. doi: 10.1037/h0020021
- Eysenck, H.J. (1968). *The scientific study of personality*. London, UK: Routledge & K. Paul.
- Feather, N. T. (1962). The study of persistence. *Psychological Bulletin*, 59(2), 94-115. doi: 10.1037/h0042645
- Fisch, R. (1970). *Konfliktmotivation und Examen [Conflict motivation and examinations]*. Meisenheim, Germany: A. Hain.
- French, E.G (1955). Some characteristics of achievement motivation. *Journal of Experimental Psychology*, 50(4), 232-236. doi: 10.1037/h0041764
- Fried-Buchalter, S. (1997). Fear of success, fear of failure, and the imposter phenomenon among male and female marketing managers. *Sex Roles*, 37(11-12), 847-859. doi: 10.1007/BF02936343
- Gelbort, K. R., & Winer, J. L. (1985). Fear of success and fear of failure: A multitrait-multimethod validation study. *Journal of Personality and Social Psychology*, 48(4), 1009-1014. doi: 10.1037/0022-3514.48.4.1009
- Glaser, B., & Strauss, A. (1967). *The discovery grounded theory: strategies for qualitative inquiry*. Chicago, IL Aldine.
- Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, 38(4), 404-411. doi: 10.1037/h0059831

- Gruber, N., & Kreuzpointner, L. (2013). Measuring the Reliability of Picture Story Exercises like the TAT. *PloS one*, 8(11). doi: 10.1371/journal.pone.0079450
- Haber, R. N., & Alpert, R. (1958). The role of situation and picture cues in projective measurement of the achievement motive. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society* (pp. 644-663). New York, NY: Van Nostrand.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate Data Analysis* (7th, Global ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Harackiewicz, J. M., Barron, K. E., Tauer, J. M., Carter, S. M., & Elliot, A. J. (2000). Short-term and long-term consequences of achievement goals: Predicting interest and performance over time. *Journal of Educational Psychology*, 92(2), 316-330. doi: 10.1037/0022-0663.92.2.316
- Heckhausen, H. (1963a). Eine Rahmentheorie der Motivation in zehn Thesen. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 10, 604-626.
- Heckhausen, H. (1963b). *Hoffnung und Furcht in der Leistungsmotivation*. Meisenheim, Germany: Anton Hain.
- Heckhausen, H. (1980). *Motivation und handeln*. New York, NY: Springer-Verlag.
- Heckhausen, H., & Halisch, F. (1986). Operant" versus "respondent" motive measures: A problem of validity or of construct. *Munich, FRG: Max-Planck-Institut für Psychologische Forschung*.
- Heckhausen, H., Schmalt, H.-D., & Schneider, K. (1985). *Achievement motivation in perspective*. New York, NY: Academic Press.
- Henry, W. (1956). *The analysis of fantasy: The Thematic Apperception Technique in the study of personality*. New York: Wiley.

- Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist*, 52(12), 1280-1300. doi: 10.1037/0003-066X.52.12.1280
- Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 30, pp. 1-46). New York, NY: Academic Press.
- Hofer, J. (2010). Research on Implicit Motives Across Cultures. In O. C. Schultheiss & J. C. Brunstein (Eds.), *Implicit motives* (pp. 433-467). New York, NY: Oxford University Press.
- Hofer, J., & Chasiotis, A. (2004). Methodological considerations of applying a tat-type picture-story test in cross-cultural research. *Journal of Cross-Cultural Psychology*, 35(2), 224-241. doi: 10.1177/0022022103262246
- Hofer, J., Chasiotis, A., Friedlmeier, W., Busch, H., & Campos, D. (2005). The Measurement of Implicit Motives in Three Cultures. *Journal of Cross-Cultural Psychology*, 36(6), 689-716. doi: 10.1177/0022022105280510
- Inquisit 3.0.6.0 [Computer software]. (2010). Seattle, WA: Millisecond Software
- Jackson, D. (1974). *Personality research form manual*. Goshen, NY: Research Psychologists Press.
- Jacobs, B. (1958). A method for investigating the cue characteristics of pictures. In J. W. Atkinson (Ed.), *Motives in Fantasy, Action, and Society* (pp. 617-629). Princeton, NJ: Van Nostrand.
- James, W. (1890). *The principles of psychology, Vol I*. New York, NY: Henry Holt and Co.

- Job, V., & Brandstätter, V. (2009). Get a taste of your goals: Promoting motive goal congruence through affect-focus goal fantasy. *Journal of Personality*, 77(5), 1527-1559. doi: 10.1111/j.1467-6494.2009.00591.x
- Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data: an introduction to cluster analysis*. Hoboken, NJ: John Wiley & Sons, Inc.
- Kehr, H. M. (2004). Integrating implicit motives, explicit motives, and perceived abilities: The compensatory model of work motivation and volition. *Academy of management review*, 29(3), 479-499. doi: 10.5465/AMR.2004.13670963
- Kerr, M. K., & Churchill, G. A. (2001). Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments. *Proceedings of the National Academy of Sciences*, 98(16), 8961-8965. doi: 10.1073/pnas.161273698
- Koestner, R., Weinberger, D. R., McClelland, D. C., & Healy, J. (1988). *How motives and values interact with task and social incentives to affect performance*. Unpublished manuscript, Department of Psychology, Boston University, Boston, MA.
- Kuhl, J. (2008). Individual differences in self-regulation. In J. Heckhausen & H. Heckhausen (Eds.), *Motivation and action* (pp. 296-322). New York, NY: Cambridge University Press.
- Kuhl, J., & Scheffer, D. (1999). *Manual for scoring the Operant Motive Test (OMT)*. University of Osnabrück.
- Langan-Fox, J., & Grant, S. (2006). The thematic apperception test: Toward a standard measure of the big three motives. *Journal of Personality Assessment*, 87(3), 277-291. doi: 10.1207/s15327752jpa8703_09

- Langens, T. A., & Mörth, S. (2003). Repressive coping and the use of passive and active coping strategies. *Personality and individual differences*, 35(2), 461-473. doi: 10.1016/S0191-8869(02)00207-6
- Lejuez, C., Aklin, W. M., Jones, H. A., Richards, J. B., Strong, D. R., Kahler, C. W., & Read, J. P. (2003). The balloon analogue risk task (BART) differentiates smokers and nonsmokers. *Experimental and Clinical Psychopharmacology*, 11(1), 26-33. doi: 10.1037/1064-1297.11.1.26
- Lejuez, C., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., . . . Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, 8(2), 75-84. doi: 10.1037/1076-898X.8.2.75
- Lowell, E.L. (1952). The effect of need for achievement on learning and speed of performance. *The Journal of Psychology*, 33(1), 31-40. doi: 10.1080/00223980.1952.9712815
- Lundy, A. (1988). Instructional set and thematic apperception test validity. *Journal of Personality Assessment*, 52(2), 309-320. doi: 10.1207/s15327752jpa5202_12
- Massey, D. S. (2008). *New faces in new places: The changing geography of American immigration*. New York, NY: Russell Sage Foundation.
- McAdams, D. P. (1980). A thematic coding system for the intimacy motive. *Journal of Research in Personality*, 14(4), 413-432. doi: 10.1016/0092-6566(80)90001-X
- McClelland, D.C. (1942). Functional autonomy of motives as an extinction phenomenon. *Psychological Review*, 49(3), 272-283. doi: 10.1037/h0054303

- McClelland, D. C. (1965). N achievement and entrepreneurship: A longitudinal study. *Journal of Personality and Social Psychology, 1*(4), 389-392. doi: 10.1037/h0021956
- McClelland, D. C. (1975). *Power: The inner experience*. New York, NY: Irvington.
- McClelland, D. C. (1987). *Human motivation*. New York, NY: Cambridge University Press.
- McClelland, D.C., & Atkinson, J.W. (1948). The projective expression of needs: I. The effect of different intensities of the hunger drive on perception. *The Journal of Psychology, 25*(2), 205-232. doi: 10.1080/00223980.1948.9917371
- McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. (1953). *The achievement motive*. New York, NY: Appleton-Century-Crofts.
- McClelland, D.C., & Boyatzis, R.E. (1982). Leadership motive pattern and long-term success in management. *Journal of Applied psychology, 67*(6), 737-743. doi: 10.1037/0021-9010.67.6.737
- McClelland, D.C., Clark, R.A., Roby, T.B., & Atkinson, J.W. (1949). The projective expression of needs: IV. The effect of the need for achievement on thematic apperception. *Journal of Experimental Psychology, 39*(2), 242-255. doi: 10.1037/h0062969
- McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How Do Self-Attributed and Implicit Motives Differ? *Psychological Review, 96*(4), 690-702. doi: 10.1037/0033-295X.96.4.690

- McClelland, D.C., & Lieberman, A.M. (1949). The effect of need for achievement on recognition of need-related words. *Journal of Personality*, 18(2), 236-251. doi: 10.1111/j.
- Michotte, A., & Prüm, E. (1910). Étude expérimentale sur le choix volontaire et ses antécédents immédiats. *Archives de Psychologie*, 12, 1-13.
- Miller, J. G. (1939). Discrimination without awareness. *The American Journal of Psychology*, 52(4), 562-578.
- Morgan, C. D., & Murray, H. A. (1935). A method for examining fantasies: The Thematic Apperception Test. *Archives of Neurology and Psychiatry*, 34, 289-306.
- Moulton, R. W. (1958). Notes for a projective measure of fear of failure. In J. W. Atkinson (Ed.), *Motives in fantasy action and society* (pp. 563-571). Princeton, NJ: Van Nostrand.
- Murray, H. A. (1933). The effect of fear upon estimates of the maliciousness of other personalities. *The Journal of Social Psychology*, 4(3), 310-329.
- Murray, H. A. (1938). *Explorations in personality: a clinical and experimental study of fifty men of college age*. New York, NY: Oxford University Press.
- Murray, H.A. (1943). *The Thematic Apperception Test: Manual*. Cambridge, MA: Harvard University Press.
- Murstein, B. I. (1965). New thoughts about ambiguity and the TAT. *Journal of Projective Techniques and Personality Assessment*, 29(2), 219-225. doi: 10.1080/0091651X.1965.10120199

- Murstein, B. I. (1965a). Projection of hostility on the TAT as a function of stimulus, background, and personality variables. *Journal of Consulting Psychology*, 29(1), 43-48. doi: 10.1037/h0021666
- Murstein, B. I. (1965b). Reliability. In B. I. Murstein (Ed.), *Handbook of projective techniques* (pp. 189-218). New York, NY: Basic Books.
- Murstein, B. I. (1972). Normative written TAT responses for a college sample. *Journal of Personality Assessment*, 36(2), 109-147. doi: 10.1080/00223891.1972.10119739
- Mücher, H., & Heckhausen, H. (1962). Influence of mental activity and achievement motivation on skeletal muscle tonus. *Perceptual and Motor Skills*, 14, 217-218.
- Norušis, M. (2008). *SPSS 16.0 statistical procedures companion* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Oertig, D., Schüler, J., Schnelle, J., Brandstätter, V., Roskes, M., & Elliot, A. J. (2013). Avoidance goal pursuit depletes self-regulatory resources. *Journal of Personality*, 81(4), 365-375. doi: 10.1111/jopy.12019
- Orso, D. P. (1969). Comparison of achievement and affiliation arousal on nAch. *Journal of Projective Techniques and Personality Assessment*, 33(3), 230-233. doi: 10.1080/0091651X.1969.10120589
- Pang, J. S. (2006). *A revised content-coding measure for hope of success (HS) and fear of failure (FF)*. Unpublished Dissertation, University of Michigan, Ann Arbor, MI.
- Pang, J. S. (2010a). The Achievement Motive: A Review of Theory and Assessment of n Achievement, Hope of Success, and Fear of Failure. In O. C. Schultheiss & J. C. Brunstein (Eds.), *Implicit motives* (pp. 30-71). New York, NY: Oxford University Press.

- Pang, J. S. (2010b). Content Coding Methods in Implicit Motive Assessment: Standards of Measurement and Best Practices for the Picture Story Exercise. In O. C. Schultheiss & J. C. Brunstein (Eds.), *Implicit motives* (pp. 119-151). New York, NY: Oxford University Press.
- Pang, J. S., Cheng, Y.-C., & Lim, H. (2008). *A PSE picture set for measuring achievement—lessons from USA and Singapore*. Paper presented at the The 22nd Congress of the International Association for Cross Cultural Psychology, Bremen, Germany.
- Pang, J. S., & Schultheiss, O. C. (2005). Assessing implicit motives in US college students: Effects of picture type and position, gender and ethnicity, and cross-cultural comparisons. *Journal of Personality Assessment*, 85(3), 280-294. doi: 10.1207/s15327752jpa8503_04
- Pang, J. S., Villacorta, M. A., Chin, Y. S., & Morrison, F. J. (2009). Achievement motivation in the social context: Implicit and explicit Hope of Success and Fear of Failure predict memory for and liking of successful and unsuccessful peers. *Journal of Research in Personality*, 43(6), 1040-1052. doi: 10.1016/j.jrp.2009.08.003
- Patten, R. L., & White, L. A. (1977). Independent effects of achievement motivation and overt attribution on achievement behavior. *Motivation and Emotion*, 1(1), 39-59. doi: 10.1007/BF00997580
- Puca, R. M., & Schmalt, H.-D. (1999). Task enjoyment: A mediator between achievement motives and performance. *Motivation and Emotion*, 23(1), 15-29. doi: 10.1023/A:1021327300925

- Qualtrics [Computer software]. (2011-2013). Provo, UT: Qualtrics
- Raphelson, A.C. (1957). The relationships among imaginative, direct verbal, and physiological measures of anxiety in an achievement situation. *The Journal of Abnormal and Social Psychology*, 54(1), 13-18. doi: 10.1037/h0041374
- Rasmussen, E. M. (1992). Clustering Algorithms. In W. B. Frakes & R. Baeza-Yates (Eds.), *Information Retrieval: Data Structures & Algorithms* (pp. 419-442). Upper Saddle River, NJ: Prentice Hall.
- Raynor, J. O., Entin, E. E., & Brown, E. T. (1982). *Motivation, career striving, and aging*. Washington, DC: Hemisphere Pub. Co.
- Reitman, W. R., & Atkinson, J. W. (1958). Some methodological problems in the use of thematic apperceptive measures of human motives. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society* (pp. 664-683). Princeton, NJ: Van Nostrand.
- Romesburg, H. C. (2004). *Cluster analysis for researchers*. Raleigh, NC: Lulu.
- Roney, C. J., Higgins, E. T., & Shah, J. (1995). Goals and framing: How outcome focus influences motivation and emotion. *Personality and Social Psychology Bulletin*, 21(11), 1151-1160. doi: 10.1177/01461672952111003
- Roth, S., & Cohen, L. J. (1986). Approach, avoidance, and coping with stress. *American Psychologist*, 41(7), 813-819. doi: 10.1037/0003-066X.41.7.813
- Sanford, R.N., Adkins, M.M., Miller, R.B., & Cobb, E.A. (1943). Physique, personality and scholarship: a cooperative study of school children. *Monographs of the Society for Research in Child Development*, 8, 1-705. doi: 10.2307/1165424

- Sarason, S. B., & Mandler, G. (1952). Some correlates of test anxiety. *The Journal of Abnormal and Social Psychology*, 47(4), 810-817. doi: 10.1037/h0060009
- Scheffer, D. (2000). *Entwicklungsbedingungen impliziter Motive: Bindung, Leistung & Macht [Developmental conditions of implicit motives: Affiliation, achievement, and power]*. Unpublished doctoral dissertation, University of Osnabrück. Osnabrück, Germany.
- Scheffer, D., Kuhl, J., & Eichstaedt, J. (2003). Der Operante Motiv-Test (OMT): Inhaltsklassen, Auswertung, psychometrische Kennwerte und Validierung (The operant motive test (OMT): Contents, scoring, psychometric values, and validation). In F. Rheinberg & J. Stiensmeier-Pelster (Eds.), *Diagnostik von Motivation und Selbstkonzept [Diagnostic of motivation and self-concept]* (pp. 151-167). Göttingen, Germany: Hogrefe.
- Schneider, K. (1978). Atkinson's "risk preference" model: Should it be revised? *Motivation and Emotion*, 2(4), 333-344. doi: 1007/BF00993328
- Schuler, H., Thornton, G. C., Frintrup, A., & Mueller-Hanson, R. (2004). *Achievement motivation inventory manual*. Oxford, UK: Hogrefe & Huber.
- Schüler, J., Brandstätter, V., & Sheldon, K. M. (2013). Do implicit motives and basic psychological needs interact to predict well-being and flow? Testing a universal hypothesis and a matching hypothesis. *Motivation and Emotion*, 37(3), 480-495. doi: 10.1007/s11031-012-9317-2
- Schüler, J., Wegner, M., & Brandstätter, V. (2013). *A critical view on implicit and explicit motive measures - What do they have in common and what differentiates them?*

Paper presented at the 6th Annual Meeting of the Society for the Study of Motivation (SSM), Washington, DC.

Schultheiss, O. C. (2001). *Manual for the assessment of hope of success and fear of failure (English translation of Heckhausen's need Achievement measure)*.

Unpublished manuscript, University of Michigan. Ann Arbor, MI.

Schultheiss, O. C., & Brunstein, J. C. (2001). Assessment of implicit motives with a research version of the TAT: Picture profiles, gender differences, and relations to other personality measures. *Journal of Personality Assessment*, 77(1), 71-86. doi: 10.1207/S15327752JPA7701_05

Schultheiss, O.C., & Brunstein, J.C. (2005). An implicit motive perspective on competence. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 31-51). New York, NY: Guilford.

Schultheiss, O. C., & Brunstein, J. C. (2010). *Implicit motives*. New York, NY: Oxford University Press.

Schultheiss, O. C., & Murray, T. (2002). *Hope of success/fear of failure questionnaire*. Unpublished manuscript, University of Michigan. Ann Arbor, MI.

Schultheiss, O. C., & Pang, J. S. (2007). Measuring implicit motives. In R. Robins, R. Fraley & R. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 322-344). New York: The Guildford Press.

Schultheiss, O. C., & Rohde, W. (2002). Implicit power motivation predicts men's testosterone changes and implicit learning in a contest situation. *Hormones and Behavior*, 41(2), 195-202. doi: 10.1006/hbeh.2001.1745

- Scott, W. A. (1956). The avoidance of threatening material in imaginative behavior. *The Journal of Abnormal and Social Psychology*, 52(3), 338-346. doi: 10.1037/h0041514
- Smith, C. (1992). *Motivation and personality: Handbook of thematic content analysis*. New York, NY: Cambridge University Press.
- Smith, C., Feld, S., & Franz, C. (1992). Methodological considerations: Steps in research employing content analysis systems. In C. Smith (Ed.), *Motivation and personality: Handbook of thematic content analysis* (pp. 515-536). New York, NY: Cambridge University Press.
- Sokolowski, K., & Kehr, H. M. (1999). Zum differentiellen Einfluß von Motiven auf die Wirkungen von Führungstrainings (MbO). *Zeitschrift für Differentielle und Diagnostische Psychologie*, 20(3), 192-202. doi: 10.1024//0170-1789.20.3.192
- Sokolowski, K., Schmalt, H.-D., Langens, T. A., & Puca, R. M. (2000). Assessing achievement, affiliation, and power motives all at once: The Multi-Motive Grid (MMG). *Journal of Personality Assessment*, 74(1), 126-145. doi: 10.1207/S15327752JPA740109
- Spangler, W. D. (1992). Validity of questionnaire and TAT measures of need for achievement: Two meta-analyses. *Psychological Bulletin*, 112(1), 140-154. doi: 10.1037/0033-2909.112.1.140
- Stanton, S. J., & Schultheiss, O. C. (2007). Basal and dynamic relationships between implicit power motivation and estradiol in women. *Hormones and Behavior*, 52(5), 571-580. doi: 10.1016/j.yhbeh.2007.07.002

- Suzuki, R., & Shimodaira, H. (2006). Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 22(12), 1540-1542. doi: 10.1093/bioinformatics/btl117
- Suzuki, R., & Shimodaira, H. (2011). Pvcust: an R package for assessing the uncertainty in hierarchical clustering (Version 1.2-2). Retrieved from <http://cran.r-project.org/web/packages/pvcust/index.html>
- Teevan, R. C., & Stamps, L. W. (1973). A motivational correlate of Viet Nam protest group members. *Psychological Reports*, 33(3), 777-778.
- Thrash, T. M., Elliot, A. J., & Schultheiss, O. C. (2007). Methodological and dispositional predictors of congruence between implicit and explicit need for achievement. *Personality and Social Psychology Bulletin*, 33(7), 961-974. doi: 10.1177/0146167207301018
- Thrash, T. M., & Hurst, A. L. (2008). Approach and avoidance motivation in the achievement domain: Integrating the achievement motive and achievement goal traditions. In A. J. Elliot (Ed.), *Handbook of approach and avoidance motivation* (pp. 217-233). New York, NY: Psychology Press.
- Tran, X., & Philipp, S. F. (2010). Financial Performance and Vision Statements of National Tourism Organizations in Northeast Asia. *E-Review of Tourism Research*, 8(3), 57-70.
- Uleman, J. S. (1972). The need for influence: Development and validation of a measure, and comparison with the need for power. *Genetic Psychology Monographs*, 85(2), 157-214.

- van de Vijver, F. J., & Leung, K. (2000). Methodological issues in psychological research on culture. *Journal of Cross-Cultural Psychology*, 31(1), 33-51. doi: 10.1177/0022022100031001004
- Vane, J. R. (1981). The thematic apperception test: A review. *Clinical Psychology Review*, 1(3), 319-336. doi: 10.1016/0272-7358(81)90009-X
- Veroff, J. (1957). Development and validation of a projective measure of power motivation. *The Journal of Abnormal and Social Psychology*, 54(1), 1-8. doi: 10.1037/h0044832
- Veroff, J. (1982). *Assertive motivations: Achievement versus power*. San Francisco, CA: Jossey-Bass.
- Veroff, J., Atkinson, J. W., Feld, S. C., & Gurin, G. (1960). The use of thematic apperception to assess motivation in a nationwide interview study. *Psychological Monographs: General and Applied*, 74(12), 1-32. doi: 10.1037/h0093760
- Weinberger, J., Cotler, T., & Fishman, D. (2010). The duality of affiliative motivation. In O. C. Schultheiss & J. C. Brunstein (Eds.), *Implicit motives* (pp. 71-89). New York, NY: Oxford University Press.
- Weiner, B. (2013). Little-known truths, quirky anecdotes, seething scandals, and even some science in the history of (primarily achievement) motivation. *Personality and Social Psychology Review*, 17(3), 293-304. doi: 10.1177/1088868313484587
- Wendt, H.W. (1955). Motivation, effort, and performance. In D. C. McClelland (Ed.), *Studies in motivation* (pp. 448-459). New York, NY: Appleton-Century-Crofts.
- Winter, D. G. (1973). *The power motive*. New York, NY: The Free Press.

- Winter, D. G. (1989). *Manual for scoring motive imagery in running text*. Unpublished manuscript, University of Michigan. Ann Arbor, MI.
- Winter, D. G. (1991). Measuring personality at a distance: Development of an integrated system for scoring motives in running text. In A. J. Stewart, J. M. Healy & D. J. Ozer (Eds.), *Perspectives in personality: Approaches to understanding lives* (pp. 59-89). London, UK: Jessica Kingsley Publishers.
- Winter, D. G. (1994). *Manual for scoring motive imagery in running text*. Unpublished manuscript. University of Michigan. Ann Arbor, MI.
- Winter, D. G. (2002). Motivation and political leadership. In L. Valenty & O. Feldman (Eds.), *Political leadership for the new century: Personality and behavior among American leaders* (pp. 25-47). Westport, CT: Praeger.
- Winter, D. G. (2005). Things I've learned about personality from studying political leaders at a distance. *Journal of Personality*, 73(3), 557-584. doi: 10.1111/j.1467-6494.2005.00321.x
- Winter, D. G. (2010). Political and historical consequences of implicit motives. In O. C. Schultheiss & J. C. Brunstein (Eds.), *Implicit motives* (pp. 407-432). New York, NY: Oxford University Press.
- Winter, D. G., & Stewart, A. J. (1977). Content analysis as a technique for assessing political leaders. In M. G. Hermann (Ed.), *A psychological examination of political leaders* (pp. 28-61). New York, NY: Free Press.
- Woike, B. A. (2008). A functional framework for the influence of implicit and explicit motives on autobiographical memory. *Personality and Social Psychology Review*, 12(2), 99-117. doi: 10.1177/1088868308315701

- Woike, B. A., & McAdams, D. P. (2001). TAT-based personality measures have considerable validity. *American Psychological Society Observer*, 14(5).
- Worchel, F. T., Aaron, L. L., & Yates, D. F. (1990). Gender bias on the thematic apperception test. *Journal of Personality Assessment*, 55(3-4), 593-602. doi: 10.1080/00223891.1990.9674093
- Wyatt, F. (1947). The interpretation of the Thematic Apperception Test. *Rorschach Research Exchange and Journal of Projective Techniques*, 11(1), 21-25. doi: 10.1080/10683402.1947.10381524
- Yaghini, M. (2010). Types of data in cluster analysis Retrieved January 20th, 2014, from http://webpages.iust.ac.ir/yaghini/Courses/Data_Mining_881/DM_05_01_Introduction.pdf

Appendix A

Constituent images for the ambiguous picture set

L-R: Basketball, Blueprints

L-R: Lecture Theater, Skaters

L-R: Operating Theater, Student

L-R: Hooligan Attack, Couple by River

Appendix B

Constituent images for the unambiguous picture set

L-R: Basketball, Blueprints

L-R: Forehead, Lecture Theater

L-R: Operating Theater, Skaters

L-R: Squash, Student

Appendix C

Constituent images for the comparison picture set

L-R: Chemist, Director, Gymnast

L-R: Man at Desk, Piano Lesson, Ship Captain

L-R: Soccer Duel, Women in Lab

Appendix D

Simplified Manual for the Assessment of Approach- and Avoidance-Related Achievement Motivation

Jonathan E. Ramsay and Joyce S. Pang

Nanyang Technological University

Singapore

Coding Guidelines

Hope of Success (HS)	Fear of Failure (FF)
1. Need for Success Scored whenever an individual sets a positively-framed achievement goal, or feels impelled to pursue such a goal.	1. Need to Avoid Failure Scored whenever an individual sets a negatively-framed achievement goal, or feels impelled to pursue such a goal.
2. Expectation of Success Scored whenever an individual expects themselves, or others, to succeed in reaching an achievement goal.	2. Expectation of Failure Scored whenever an individual expects themselves, or others, to fail to reach an achievement goal.
3. Positive Affect Scored whenever an individual experiences positive affect as a result of reaching, or progressing towards, an achievement goal.	3. Negative Affect Scored whenever an individual experiences negative affect as a result of failing to reach, or failing to progress towards, an achievement goal.
4. Success Scored whenever an individual succeeds in reaching an achievement goal, or attains a standard of excellence.	4. Failure Scored whenever an individual fails to reach an achievement goal, or falls short of a standard of excellence.

GENERAL RULES

What Constitutes Motive Imagery?

1. Achievement imagery is scored for actions or internal states (e.g., needs, emotions) that arise in pursuit of an achievement goal. Reference to an achievement goal, either explicit or strongly implied, is a necessary prerequisite for scoring achievement imagery.
2. Such actions and internal states are scored regardless of their origin or source. Whether they are attributed to an individual, a group, or to people in general, all such instances are scored. Self-attributions made by the speaker, author, or narrator are also scored.
3. Achievement imagery that is negated is not scored.
4. Achievement imagery that is described as uncertain or possible is scored as if it were definite. Only full negations (see point 3) are not scored.

Identifying True Achievement Motive Imagery

5. True achievement motive imagery is indicated by the presence of a standard of excellence. Such standards can be personal (e.g., concern with improvement, mastery) or interpersonal (e.g., concern with competition, performance).
6. To be scored as achievement motive imagery, relevant actions or internal states must be self-endorsed and non-trivial, two criteria that indicate the presence of a standard of excellence.
7. Emotional involvement in an achievement task indicates the presence of an implicit standard of excellence, and associated internal states and actions should be scored.
8. However, if achievement-related actions are performed only in order to satisfy external demands (e.g., occupational duties, requests from others, feelings of obligation), achievement motive imagery is not scored.
9. Similarly, if achievement-related actions are described as being trivial, routine, or personally unimportant to the actor, achievement motive imagery is not scored.
10. Use of cheating or deception negates further scoring of achievement imagery. Actions or internal states that precede the deception can be scored, but all subsequent actions or internal states (relating to the activity where cheating has been employed) should be ignored.
11. Achievement imagery that is later downplayed or belittled, or revealed to be a joke or farce, is discounted and any scores should be removed.

12. Achievement goals that are set or pursued in order to satisfy other motives, such as power or affiliation, are not scored. Conversely, power and affiliation goals set or pursued in order to satisfy an underlying achievement need are scored as achievement imagery.
13. If imagery relates to both need for achievement and an additional motive (e.g., nAff or nPow), and neither is clearly prioritized, score for achievement imagery.
14. Goals reflecting disinterest or a preference for other activities are not scored for HS.

What are the Different Forms of Achievement Motive Imagery?

15. Achievement imagery exists in two forms – Hope of Success (HS) and Fear of Failure (FF).
16. HS-type achievement imagery is scored for actions or internal states that arise through positive, promotion-focused pursuit of a desired outcome or end-state.
17. FF-type achievement imagery is scored for actions or internal states that arise through negative, prevention-focused attempts to avoid an undesired outcome or end-state.
18. Both HS and FF imagery can manifest in different forms, corresponding to different actions and internal states. See the coding guidelines on page 199 for a description of these different forms.

How Can I Calculate the Amount of Motive Imagery in a Piece of Text?

- a) When scoring for achievement imagery, the sentence is the unit of analysis, and text should be analyzed on a sentence-by-sentence basis.
- b) HS and FF are scored simultaneously. For each sentence, the coder must identify whether HS or FF imagery are present.
 1. If HS imagery is present, the coder scores for HS.
 2. If FF imagery is present, the coder scores for FF.
 3. If both HS and FF imagery are present, the coder scores both HS and FF.
 4. If neither HS nor FF imagery are present, the coder scores nothing.
- c) Each sentence can only be scored once for HS and once for FF
- d) Once every sentence in the text has been analyzed, the coder should add up the scores for HS and FF separately, yielding two raw motive scores.
- e) If the text is split into several sections (as in responses to a PSE picture set), the raw motive scores should then be summed. This yields two aggregate raw motive scores: one for HS and one for FF.

- f) Each of these aggregate scores should then be divided by the total number of words in the piece(s) of text, and then multiplied by 1000. This calculates the number of instances of motive imagery per thousand words, and these word-count corrected HS and FF scores should be used in subsequent analyses to avoid the potential confound of verbosity.

HS1: NEED FOR SUCCESS

Definition:

- Scored whenever an individual sets a positively-framed achievement goal, or feels impelled to pursue such a goal.

Additional Information:

- Oftentimes these positively-framed achievement goals are stated in terms of desires, such as “wanting” or “needing” to reach a certain positive achievement outcome, or “wishing” or “hoping” that the positive outcome will be achieved.
- In accordance with the general rules above, the positively-framed achievement goal must involve reaching a standard of excellence, must be self-endorsed and non-trivial in nature. As such, positively-framed achievement goals that arise entirely due to external demands (i.e., they are not self-endorsed), and those that involve completion of an easy or routine task (i.e., they are trivial), are not scored for HS.
- Occasionally it may be stated that an individual feels they “have to” or “must” reach a certain positively-framed achievement goal, in order to meet external expectations or demands. This should only be scored for HS as long as it is clear that they have internalized these demands (i.e., the achievement goal is pursued in a self-endorsed fashion).
- Negatively-framed achievement goals (i.e., those focused on avoiding failure), are not scored for HS. Rather, they qualify as avoidant imagery and are scored for FF (see page 209).

Examples:

Scored for HS

- The soccer player wants to win the match
- “I really hope I get a good grade on this project”, thinks the student
- The businessman begins his presentation, hoping to secure new business

Not Scored for HS

- Lining up his putt, the golfer hopes he won’t miss the short (negatively-framed goal)
- The boy wants to finish his homework so he can play video games (no standard of excellence)
- The woman wants to read the document (trivial goal, no standard of excellence)

HS2: EXPECTATION OF SUCCESS

Definition:

- Scored whenever an individual expects themselves, or others, to succeed in reaching an achievement goal.

Additional Information:

- Expectations of success do not have to be certain to be scored for HS, but success must clearly be judged as being more likely than failure.
- Expectation of success must relate to positive outcomes that are anticipated to result from good performance in order to be scored as HS. Positive expectations regarding luck or the influence of external circumstances are not scored for HS.
- Expectations of praise or reward are scored for HS, so long as the anticipated praise or reward is contingent on reaching a standard of excellence.

Examples:

Scored for HS

- The basketball player tries to concentrate. “I can do this”, he thinks
- The student will probably do well in the science contest
- Contemplating his recent performance, the employee thinks he’s certain to get a raise

Not Scored for HS

- Once the experiment is complete, the researcher’s findings may or may not be published (success not seen as being more likely than failure)
- After blackmailing the judge with incriminating photographs, the skater knows he will get a good score (expected success not contingent on good performance)
- I expect the man will finish his work (trivial, no standard of excellence)

HS3: POSITIVE AFFECT

Definition:

- Scored whenever an individual experiences positive affect as a result of reaching, or progressing towards, an achievement goal – i.e., when they succeed in attaining a standard of excellence.

Additional Information:

- HS is also scored when an individual experiences positive affect as a result of *someone else* reaching or progressing towards an achievement goal.
- In order to qualify as an achievement goal, the activity must be non-trivial and engaged in autonomously (see general scoring rules above).
- Only positive affective states that RESULT from success or mastery should be scored for HS. Positive affective states that motivate or cause achievement-related activity, such as “enthusiasm” or “excitement”, are not scored for HS unless the positive affective state is itself a consequence of previous success or mastery. “Enjoyment” is another difficult case. Simple enjoyment of an achievement-related activity should not be scored for HS unless that enjoyment arises from goal progress or attainment.
- Positive affective states that result from avoidance of failure are not scored for HS. Similarly, relief is not scored for HS since it typically refers to the alleviation of a negative affective state.

Examples:

Scored for HS

- The father is proud of his daughter’s academic achievements
- Crossing the finish line in first place, the cyclist was overjoyed
- Now that his backhand is improving, the boy is enjoying his tennis lessons more

Not Scored for HS

- She was glad that the test score hadn’t compromised her perfect GPA (positive affect arises from avoidance of failure)
- The man is excited to be participating in the photography contest (positive affect does not result from achievement or mastery)
- The young designer is glad to have completed the boring project (positive affect does not arise from achieving a standard of excellence)

HS4: SUCCESS

Definition:

- Scored whenever an individual succeeds in reaching an achievement goal, or attains a standard of excellence.

Additional Information:

- Typical examples would be winning a sports competition, gaining a promotion, or getting a good grade on a term paper.
- Partial successes can also be scored for HS. Small successes that bring an individual closer to attaining a larger achievement goal are also scored.
- Successes can also take the form of unique accomplishments, such as world records, scientific discoveries, and new inventions.
- References to past successes are also scored for HS.
- Successes that result from accidents of fate or another person's intervention are not scored for HS. Success must be contingent on good performance.

Examples:

Scored for HS

- Jiamin won first prize in the archery tournament
- His last round of testing yielded encouraging results
- Frederick had always finished top of the class, but he was still nervous

Not Scored for HS

- After her opponent was disqualified, Alexa went through to the next round of the judo competition (success not contingent on good performance)
- The taxi driver successfully located the hotel (trivial, no standard of excellence - unless the hotel is explicitly said to be difficult to find)
- The politician successfully smeared his opponent (success not achievement-related)

FF1: NEED TO AVOID FAILURE

Definition:

- Scored whenever an individual sets a negatively-framed achievement goal, or feels impelled to pursue such a goal.

Additional Information:

- Negatively-framed achievement goals are frequently stated in terms of desires, such as “wanting” or “needing” to prevent failure, or “wishing” or “hoping” that failure, or the consequences of failure, can be mitigated, lessened, or avoided entirely. These consequences can also be social, meaning that the desire to avoid negative evaluation should also be scored for FF.
- The negatively-framed achievement goal must indicate concern with maintaining a standard of excellence, and must be self-endorsed and non-trivial in nature. Negatively-framed achievement goals that arise entirely due to external demands or involve completion of an easy or routine task are not scored for FF.
- It may be stated that an individual feels they “have to” or “must” reach a certain negatively-framed achievement goal in order to meet external expectations or demands. This should only be scored for FF as long as it is clear that they have internalized these demands.
- Negatively-framed achievement goals that are set or pursued in order to satisfy other motives, such as power or affiliation, are not scored. Similarly, goals reflecting disinterest or a preference for other activities are not scored for FF.
- Positively-framed achievement goals (i.e., those focused on attaining success), are not scored for FF. Rather, they qualify as approach-related imagery and are scored for HS (see page 202).

Examples:

Scored for FF

- Jack was determined not to fail the module
- The junior executive hopes her mistake won’t damage her yearly appraisal
- “If I had only trained harder...” thought the athlete, contemplating his poor finish

Not Scored for FF

- The girl hopes her boyfriend won’t leave her (goal is not achievement-related)

- With lunch fast approaching, Pablo hopes the task won't prove too difficult (not concerned with standard of excellence)
- Desperate to impress his father, the boy wants to avoid getting a bad report card (negatively-framed achievement goal, but pursued in order to satisfy external demands)

FF2: EXPECTATION OF FAILURE

Definition:

- Scored whenever an individual expects themselves, or others, to fail to reach an achievement goal.

Additional Information:

- Expectations of failure do not have to be certain to be scored for FF, but failure must clearly be judged as being more likely than success.
- Expectation of failure must relate to negative outcomes that are anticipated to result from poor performance on the part of the individual(s) in order to be scored as FF. Negative expectations regarding bad luck or unfortunate circumstances are not scored for FF.
- Expectations of criticism or punishment are scored for FF, so long as the anticipated criticism or punishment is contingent on failing to reach a standard of excellence.
- Sometimes emotional states such as apprehension or nervousness indicate an expectation of failure, especially when they are contrasted with confidence of success. This should be scored for FF under the expectation of failure criterion. However, mention of nervousness, agitation or stress in isolation is not sufficient for scoring of FF under the expectation of failure criterion.

Examples:

Scored for FF

- The team will most likely lose their next match
- No-one ever expected Christopher to achieve anything
- “I’m not skilled enough to save this patient” thought the doctor
- Some of the students are confident about the exam, while others are nervous

Not Scored for FF

- “This patient’s injuries are too severe, I don’t think I can save them” thought the doctor (expected failure due to unfortunate circumstances, does not reflect lack of excellence)
- Katie probably won’t make it to her dental appointment on time (trivial, not achievement-related)
- He might win, he might lose... I’m not sure (failure not deemed more likely than success)

- Jack is completing the exam, and feels stressed (stress not clearly indicative of expected failure, could just be generalized arousal)

FF3: NEGATIVE AFFECT

Definition:

- Scored whenever an individual experiences negative affect as a result of failing to reach, or failing to progress towards, an achievement goal (i.e. when they fail to attain a standard of excellence).

Additional Information:

- HS is also scored when an individual experiences negative affect as a result of *someone else* failing to reach or failing to progress towards an achievement goal.
- In order to qualify as an achievement goal, the activity must non-trivial and engaged in autonomously (see general scoring rules above).
- Simple dislike of an achievement-related activity should not be scored for FF unless the negative attitude arises from failure to reach or progress towards an achievement goal.

Examples:

Scored for FF

- After falling from the beam, the gymnast was inconsolable
- Unable to improve design, the architect is frustrated
- Coach Mayhew was bitterly disappointed with his team's insipid performance

Not Scored for FF

- Rebecca does not enjoy calculus (negative affect does not arise from failure)
- Vivek was upset that his exam results did not meet his father's expectations (achievement goal is not self-endorsed)
- The carpenter is annoyed that he can't find his tools (trivial, negative affect does not arise from achievement-related failure)

FF4: FAILURE

Definition:

- Scored whenever an individual fails to reach an achievement goal, or falls short of a standard of excellence.

Additional Information:

- Typical examples would be losing a sporting competition, being fired from a job, or getting a poor grade on a term paper.
- Partial failure can also be scored for FF. Unsuccessful instrumental activities that hinder progress towards an achievement goal are also scored.
- References to past failures are also scored for FF.
- Failures that result purely from bad luck or another person's negative intervention are not scored for FF. Failure must be contingent on poor performance.
- Failures that are subsequently remedied by the actor in question are discounted and should not be scored. However if the failure is remedied by someone else, FF is still scored.

Examples:

Scored for FF

- The director's previous film was a flop
- Carl has been competing for years, but has never made it past the first round
- Nadia is stuck on the final question

Not Scored for FF

- The class were given the wrong exam paper, so they all failed (failure due to external circumstances, does not reflect lack of excellence)
- Henry lost the first two sets, but went on to win the match (initial failure is remedied)
- Crystal's previous relationships had all ended in failure (failure not achievement-related)

Achieving Reliability with the System

Coders attempting to learn this system should score these practice stories first. Once this is complete, they should compare their scores with the master scores for each story (see page 223). The concordance rate between two sets of scores should be calculated as follows:

$$\frac{(\text{No. Scores by Coder} + \text{No. Master Scores})}{(2 * \text{No. Coding Agreements between Coder and Master Scores})}$$

Separate concordance rates should be calculated for HS and FF. If the concordance rates are .85 or higher, the coder is considered reliable and can proceed to code other material. If either of the concordance rates is below .85, the coder should rescore the practice stories until a concordance rate of .85 is achieved.

Practice Stories

1. There are two guys playing squash. They look like best friends to me as there are no competition going on. This can be seen from the guy behind who is just waiting for his friend to pick up the ball. They seem to be happy and relaxed while playing the game.
2. It looks like John (the person in picture) has just missed a crucial ball in a game of tennis. He is clutching his head in a sign of disappointment with himself, as it was the match point for him and he let it slip by his grasp by a near miss. However, after that brief moment of anger, he picks himself up from his sad state and does not let it affect his other games, going on to win his next three sets despite losing the first one.
3. Siblings playing in the background. In the foreground, the elder sister tries to stifle laughter and concentrate on finishing her assignment at hand. It's a lab report, which requires much precision and accuracy, failing which she would have to redo her calculations and analysis. It's almost complete and after this, she would be going out on a date with her best friend to town, where they would catch a movie and grab some dinner.
4. A student doing homework. She is thinking about the solution and wants to solve the question well. She will finish the homework, and then relax.
5. It's the end of the semester and Joseph is fighting for his first class honours as it's his last chance. However, the questions are not so simple. Joseph tried very hard and finally got the problem solved.
6. John and Wendy are sitting on a bench by the river. The bridge sits to their right and the weather seems calm. The mood is sombre as John attempts to speak to Wendy. She is distraught and has no feeling or emotion left to show. John had just shared with Wendy that he is due to leave for Iraq in a week and that he would be there for about four months. The sudden thought of losing John had been the first thing that had popped into Wendy's mind. She had no way of responding to this as she had never thought that her boyfriend would have to be

sent overseas to aid in the war effort. Now that the time had come, she had no way of truly showing how she felt. She continued to sit motionless as John tried all means to communicate with her. John did not have a clue as to how Wendy was feeling and continued to try his very best to get a response from her.

7. The setting is probably a classroom, or perhaps an examination. The girl looks like she is concentrating on solving the questions given to her. She is probably cracking her brains to get the answer. Looking at the seriousness on her face, she is probably looking to score as high as possible for this examination paper. Time will be up soon, then they will hand in the examination paper.
8. A pretty girl named Ifa is sitting for an exam. One question in the exam is not easy, even though she usually practices one month before. She tries to think about the problem that she cannot do until now. All the people around Ifa seem like they are having the same problem, and most likely they will see each other and reflect their emotion.
9. The doctors are about to perform a surgery. They could be thinking about a lot of things. After all, they have been doctors for a long time and surgeries like this one could be just another day at the hospital for them. I hope they are not thinking about their kids, or the bills they need to pay. I hope for the sake of the patient that they are fully concentrated on the task at hand. I think doctors are responsible people though, I'm sure they will do their best to save the man. The man lives a long and healthy life.
10. Michael, the guy with the bandana wrapped around his forehead, is the star of the team. This scene is taking place during basketball training. The coach has them play a match where a group of defenders are supposed to defend Michael from scoring single-handedly. Being surrounded by his team mates, Michael is surrounded and trapped. He tries to find an alternative solution. He keeps his focus on the net and strategizes a lay-up. His team mates feel challenged knowing of Michael's talents. Yet they try their best. Nonetheless, Michael's strategy outwits them as in a series of sly tricks he breaks out from the enclosed space he was in to score a shot.
11. A small surgery is taking place. I think a doctor and his nurse are in charge of the surgery. The doctor is focusing on a pair of tweezers with a thread-like specimen on it. He should be thinking about how to stitch a wound. The nurse seems to be focusing on something else as she is not standing close to the doctor. The doctor should be feeling as per normal as he should not let his emotions affect the surgery. I think they want to help the patient. The doctor will praise the nurse after the surgery for a good job done.
12. Sam and Jean are ice skaters for the school team, they have been practising hard for the state finals which are coming up next week. This is the most important competition of their life as the school has never found a championship before. With winning in their mind, they practised day and night with little rest, sometimes even skipping meals. Two days before the competition, the rival school appears and challenges them to a skate-off. Thinking that this

would be a good opportunity to practise, they agreed. However, they did not realise that their skates were tampered with. As the duo were about to perform their trademark move, The Flying Swan. Jean's skates fell apart and she crashed to the ground, fracturing her leg.

13. The picture shows an ice skating rink where two people are in competition. Two sportswomen in action. The woman in the background is playing catch-up as she is closing up the leading woman. The woman in the lead appears confident of her chances of winning. They both want to win. The woman in the lead will lose her lead to overconfidence and complacency. They will receive their medals after the championship and celebrate together.
14. A young man and lady sit on a wooden bench by the river bridge. The man stares at the girl with questioning eyes. The girl has told the man that she has to leave the city with her family and can no longer be with him. The man is filled with disbelief and asks, "Is there nothing I can do?" The lady shakes her head and a tear escapes her eyelid. She stands up and starts to walk away but the man takes her hand. "Let's run! To a place far away from here!", exclaims the man.
15. It was a hot and sunny day at the court. Adam was against Steven in that match and it was the deciding shot for the match. Adam served a shot to Steven and as he attempted to return, he slid and tumbled over on the court. He missed the shot. Adam won the match. As Steven sat up, he held his hand to his head in defeat. He kept thinking, "Why did I trip?" The match ended and Steven lost. He walked off the court in defeat and shame.
16. Why is it so hard to do??!! The student felt, a sense of frustration and irritation for not being able to complete a simple exercise. Always an average student, she has yet to experience any major achievement in the academic sense. Feeling the pressure from others, and worst of all herself, each passing day is a stressful one... with much difficulty and help from friends, she finally be able to solve this question and move on to the other question.
17. He is stressed out by his work. This is common in the 21st century workplace, as bosses frequently demand more than one's ability can supply. He looks very depressed and will probably burn out if he does not find an appropriate remedy for his crisis.
18. These people are in their lecture theatre sitting in on a lecture after they had their lunch break. It's two in the afternoon and they are feeling full and happy after a good lunch. But once the lecture starts, the students scramble to keep up with their professor as he is explaining the topic to them. It is difficult for them to understand as the topic is very difficult and they are getting frustrated and stressed out. They are thinking that they will have to go back and revise to be able to keep up with the rest of the class. So, they rush home after school to study.
19. Four people are playing a basketball match. The player in the centre is clearly thinking about taking the shot, disregarding his teammate who is standing right behind him and potentially has no obstruction. His opponents on the other hand are so intent on blocking his move, perhaps knowing that this is the better player that they have ignored his teammate, leaving him

potentially open to make his move. The situation is tense as one team is trying their utmost to prevent the other from scoring. However, despite appearing that he has ignored his partner, it was all part of his plan and his next move was to pass the ball back, feigning a shot and so taking his other two opponents out of the equation and allowing his team mate to score.

20. A bunch of people are fighting, and one of them has killed a guy. The culprit flees as fast as possible. The rest are shocked, not knowing immediately how to respond. In the end, all of them escape, and the poor man just dies like a fly.
21. Two males playing squash. Although it is impossible to know just by looking at the picture, it seems as if the guy in the background is anticipating the shot, while the guy in the foreground is concentrating on hitting the ball. Both will be thinking on how to counter each other.
22. Two persons are ice skating. They are competing against each other to see who is the fastest. They seem very happy while doing it.
23. This is an examination hall. People are doing an exam. Everyone is concentrating on the questions given. Some of them are looking around. They might finish the exam paper in time.
24. The exam is in progress, and the students are working on it. Some of them are feeling nervous while others are confident. They will finish the exam.
25. Jeremy had stayed up all night to study for his final year exam today. Feeling sleepy and tired, he could not concentrate on his paper. He pondered hard over the question and tried to recall what he has studied last night. As the minutes pass, he failed to answer the question. Demoralised and upset, he browsed through the papers trying to find questions that he could answer. At the end of the paper, he shook his head and walked out of the exam hall slowly.
26. She is trying to do the assignment given by her teacher in the class. She will have to submit the project by the end of the lesson, and now she is searching for inspiration to write things down. The second she gets an idea, she writes down everything she can think about. After finishing the word quotas, she is planning to submit her work to the teacher and go out of the class earlier.
27. These are ice skaters. They are just concentrating hard and focusing on skating and doing their best. This might be at a competition or rehearsal of some sort. They hope to win if it is a competition. They also seem to feel tense and stress as their faces look serious. I believe that they will win.
28. It is 7pm on the clock, and he sat at his desk, rubbing his temples tiredly. He thought to himself, "How will I ever finish this heap of work that keeps piling up?" The deadline for the tender is tomorrow and he is barely done with it. Work commitments aside, his wife and

children are waiting for him to come home to celebrate his daughter's 3 year old birthday. But how could he leave with the amount of work awaiting him? Tensions were building up at home with his wife as he always returned home late from working over-time. He knew his wife would be unhappy if he did not leave the office for home very soon. The cell phone beside him started to ring. He looked at it for a moment, rejected the call and went back to his work.

29. It is yet another long night for Dave. He had been working overtime for the past two weeks, at a job he simply hated despite putting on all smiles for his colleagues and superiors all the time. His dream was to become an artist, not sit at a desk doing some stinking nine to five job which sucked the life out of him. He thought about his parents and how they had struggled to put him through his college education, his wife who has always been so proud of him for getting promoted to his current position so quickly, and his children who seemingly had unlimited wants and desires for the newest and most expensive toys. He was tired of it all, yet he could not just stop. He smiled faintly to himself as he thought about what he was going to do next. He had planned it for the past year, making sure to tie up his loose ends.
30. A surgery is going to start, or perhaps it's just finished. Or maybe it's a medical lab in a university. There is either a surgeon and his assistant, two surgeons, or a professor and a student. They want to communicate with each other. Next they will leave the room, start the surgery, or continue to talk
31. The group is having a mid-term examination in the lecture theatre. The theatre is quiet, with everyone fully focused on the examination. The main man in the picture appears to have an expression of worry. He has prepared well for the examination, but the questions do not appear to be familiar. He tries his best. He is thinking about guessing his way through the questions he is unfamiliar with. He can only do so much to better his chances of doing well in the examination. He finishes the paper early and leaves.
32. This is a class whereby the people are waiting to complete a test. Final year paper is upon them and the atmosphere is tense and only scribbling can be heard in the class. All the students are intent on completing the paper and achieving the best grade as possible. Small Clicks can be heard from the constant walking of high heeled professors and the anxious movements of the students who are trying their best to recall the massive loads of information they have tried to memorize the night before. Meanwhile those that have studied sufficiently take their time and slowly complete the paper one page at a time. Before long, the bell signalling the end of the paper sounds and the whole hall erupts into a sigh of relief. It is finally OVER!
33. A girl is doing a math exercise in class. The exercise is very easy for her and she can come up with the solution in a short time. However, the two boys behind her are discussing the questions because they cannot finish it them by themselves. They all finish the exercise in the end.

34. It's a normal day in the hospital with a couple of surgeries on going. Dr John, on the left, is a general surgeon performing a simple surgery for his patient. He cut open the patient, fixed the problem, and is now closing the patient up. His intern, Christina, watches over and assists him as much as she can. Dr John hopes that by watching and gradually helping out in the surgeries, Christina would learn more and become a great surgeon. Christina watches on and is amazed at Dr John's fine stitching skills. She aspires to be like him.
35. It was finally the final examination. John had studied very hard for it, and he had to do well. As he stared at the paper, he felt nervous and tried to calm himself down. The paper was difficult, just like everyone had expected. He looked around, and saw that his course mates were racking their brains doing the paper. He decided to focus on his own work, and started doing it. It wasn't as hard as he initially thought. He finished the paper and felt relieved.
36. The picture shows a girl writing something with her pencil, probably taking a test. In the background there are two boys. The girl is concentrating on her work, and the boys are interacting with each other. The girl is calm and collected. The girl wants to complete her work and is not distracted by external stimuli and looks focused. The teacher will come along to collect her paper.
37. A female student taking notes in a class while the two guys in the background are busy talking. The female student will be thinking of understanding what the person she is listening too saying, while the two guys are thinking about their conversation instead.
38. The surgeon's hands shook lightly, barely visible to anyone, but he knew it himself. The man on the patient's bed was his father, the man who brought him up singlehandedly and gave him his education. Now he was lying here in pain, pleading with his eyes for his son to put him out of his misery. "What if they found out?" A million thoughts raced through the surgeon's mind, and he struggled to focus on the procedure he had rehearsed so many times. He could not afford to make any mistakes.
39. This is a test happening in the lecture theatre. The guy in the centre of the picture is stumped - he does not know the answers to the foreign looking questions on his paper. He is feeling anxious and wishes that somehow, the answers would magically appear in his head. Unfortunately, that did not happen and he left the lecture theatre calculating how well he would have to score in the upcoming tests so that a B+ is still possible for this module.
40. this is an ice-skating race. the first two skaters are sprinting to the final. both of them want to get the champion. the second skater wants to overrun the first one so as to win the race.
41. Guy: "What? You are pregnant?"
Girl: "Yes, I just verified it today."
Guy: "Oh no, what are we going to do about it? Let me digest this first. You are pregnant, but we are still so young! We are still in school! I don't have the financial means to support a baby, let alone you!"

Girl: "I know, but..."

Guy: "And our parents can't know about this! My dad will kill me! No, wait, I think your mum will kill me first!"

Girl: "You are not making me feel better..."

Guy: "Listen, we have to solve this like adults. I will find a gynaecologist. You should get an abortion ok? We are still so young! Think of the restrictions of being a mother at such a young age! Listen to me ok?"

42. Two surgeons are preparing for huge operation. They have been doing operations for years and years but still each operation makes them feel nervous. They are carefully looking through all the devices that they need for operation and they strongly wish they can save another life! Time for the entry of patient. Good luck!
43. Mother and daughter take to the ice rink and show off their skills. They become like two identical figures on the ice. Sponsored by the best of companies, their success has reached a pinnacle and it seems like nothing can stop them. Anna is not happy though, as she longs to be with the love of her life. That evening she tells her mother she loves her boyfriend Hamish and will return to him in Norway. Her mother looks at her with contempt, for Hamish is Indian and therefore not a suitable match. Her mother's prejudices have haunted Anna all her life, and she has had enough. She strangles her mother with a traditional Indian sari. It's a nice touch.
44. There is always a need to improve and learn on the job. This is the motto that the Chief of Surgery Mercy West lives by day to day. Today, he is about to guide me, a first year resident, in how to carry out a proper endoscopy. I had not previously done this procedure and to my awe I had been chosen to assist him in the surgery. An hour before the surgery, he called me in to the operating theatre and asked me to prep myself for surgery. I did so, even though I found it peculiar that we were doing so pretty early. I then found out that he was going to do a dry run and a walk through of the surgery. It was through this that I got to visualise what he expected at each stage of the surgery. I was really impressed at how the Chief was so meticulous in his job and at the same time found time to teach me the proper way of doing things.
45. This is an exam scene in a lecture theatre, and it focuses on a single guy (let's call him Mike). Mike has been studying the entire night before for this quiz after choosing to party the week away, and right now in the midst of the quiz he struggles to recall what he has been studying for so frantically. He tries to piece together the information lying somewhere in his brain, and reproduce them as best as he can on the script. He hasn't had the time to shave, and his stubble shows. The questions do not make sense, the answers aren't forthcoming, and he knows right now that he should have started studying earlier. Regret does not help him however, and he resolves to study earlier next time. But for now, this quiz would most likely return a bad result, and there is only so much he can do to salvage it.
46. A dating scene by the riverside where a couple is sitting on the bench talking. There is a guy and a girl which suggest that they are probably dating. The guy is probably trying to impress the girl.

47. This is first ice skating competition organised by McDonald's, as we can see their logo on the competitor's head band. It is time for the girl's ice skating and there are only two competitors. McDonald's totally failed to make this competition successful. However, with their sportsmanlike spirit, these two women are trying their best not to be on the second place which is equals to last place. They both want to win, but Sally who is skating fastest got first place!
48. They have been in surgery for 6 hours now. The surgeon had just transplanted the liver to the patient. All was going well. Just then, the patient began having seizures as his body jerked violently. The whole operation room was thrown into a frenzy as they tried to figure out a solution. The nurses tried to gain control over his condition as the surgeon continued the procedure. When it seemed that all hope was lost, the patient suddenly showed signs of calm and returned to a normal state.
49. Amy was falling behind with her school work and her grades were dropping. Her parents were disappointed with her and told her to pull her grades up from now on. She felt irresponsible and decided to focus more in class and work harder. This picture shows her concentration and focus while doing her work in class. She knows she can do well, and she wants to do herself and her parents proud. She sits for the examination and manages to pull her grades up to become the best student in the class. Finally, her hard work has paid off.

Master Scores

1. There are two guys playing squash. They look like best friends to me as there are no competition going on. This can be seen from the guy behind who is just waiting for his friend to pick up the ball. They seem to be happy and relaxed while playing the game.

None.

No standard of excellence.

2. It looks like John (the person in picture) has just missed a crucial ball in a game of tennis. He is clutching his head in a sign of disappointment with himself, as it was the match point for him and he let it slip by his grasp by a near miss. However, after that brief moment of anger, he picks himself up from his sad state and does not let it affect his other games, going on to win his next three sets despite losing the first one.

FF, FF, FF, HS.

Previous score for failure is discounted by ultimate success. Scores for negative affect remain.

3. Siblings playing in the background. In the foreground, the elder sister tries to stifle laughter and concentrate on finishing her assignment at hand. It's a lab report, which requires much precision and accuracy, failing which she would have to redo her calculations and analysis. It's almost complete and after this, she would be going out on a date with her best friend to town, where they would catch a movie and grab some dinner.

None.

Effort relates only to avoiding distraction, no standard of excellence.

4. A student doing homework. She is thinking about the solution and wants to solve the question well. She will finish the homework, and then relax.

HS.

Merely "finishing" homework does not clearly indicate achieving a standard of excellence, so a second HS for success is not scored.

5. It's the end of the semester and Joseph is fighting for his first class honours as it's his last chance. However, the questions are not so simple. Joseph tried very hard and finally got the problem solved.

HS.

The final sentence is scored for the successful outcome.

6. John and Wendy are sitting on a bench by the river. The bridge sits to their right and the weather seems calm. The mood is sombre as John attempts to speak to Wendy. She is distraught and has no feeling or emotion left to show. John had just shared with Wendy that he is due to leave for Iraq in a week and that he would be there for about four months. The

sudden thought of losing John had been the first thing that had popped into Wendy's mind. She had no way of responding to this as she had never thought that her boyfriend would have to be sent overseas to aid in the war effort. Now that the time had come, she had no way of truly showing how she felt. She continued to sit motionless as John tried all means to communicate with her. John did not have a clue as to how Wendy was feeling and continued to try his very best to get a response from her.

None.

No achievement-related imagery.

7. The setting is probably a classroom, or perhaps an examination. The girl looks like she is concentrating on solving the questions given to her. She is probably cracking her brains to get the answer. Looking at the seriousness on her face, she is probably looking to score as high as possible for this examination paper. Time will be up soon, then they will hand in the examination paper.

HS.

Here the fourth sentence is scored under the need for success criterion.

8. A pretty girl named Ifa is sitting for an exam. One question in the exam is not easy, even though she usually practices one month before. She tries to think about the problem that she cannot do until now. All the people around Ifa seem like they are having the same problem, and most likely they will see each other and reflect their emotion.

FF, FF.

Both sentences refer to an achievement-related problem that cannot be solved (i.e., a partial failure), first for the girl Ifa and then her fellow students. Both are scored as failure.

9. The doctors are about to perform a surgery. They could be thinking about a lot of things. After all, they have been doctors for a long time and surgeries like this one could be just another day at the hospital for them. I hope they are not thinking about their kids, or the bills they need to pay. I hope for the sake of the patient that they are fully concentrated on the task at hand. I think doctors are responsible people though, I'm sure they will do their best to save the man. The man lives a long and healthy life.

HS, HS.

The early part of the story suggests that the surgery may be viewed as routine by the surgeons (and so wouldn't be scored). However, the author later makes it clear that the surgeons are responsible individuals who care about the outcome. As such, both the need for success and the successful outcome are scored.

10. Michael, the guy with the bandana wrapped around his forehead, is the star of the team. This scene is taking place during basketball training. The coach has them play a match where a group of defenders are supposed to defend Michael from scoring single-handedly. Being surrounded by his team mates, Michael is surrounded and trapped. He tries to find an alternative solution. He keeps his focus on the net and strategizes a lay-up. His team mates

feel challenged knowing of Michael's talents. Yet they try their best. Nonetheless, Michael's strategy outwits them as in a series of sly tricks he breaks out from the enclosed space he was in to score a shot.

FF, HS.

Both success (Michael) and failure (his team mates) are scored in the final sentence.

11. A small surgery is taking place. I think a doctor and his nurse are in charge of the surgery. The doctor is focusing on a pair of tweezers with a thread-like specimen on it. He should be thinking about how to stitch a wound. The nurse seems to be focusing on something else as she is not standing close to the doctor. The doctor should be feeling as per normal as he should not let his emotions affect the surgery. I think they want to help the patient. The doctor will praise the nurse after the surgery for a good job done.

HS.

Wanting to help the patient is an affiliative goal, but also represents an achievement goal for someone in the medical profession. As such, HS is scored for wanting to help the patient, since the affiliative, compassionate nature of the goal is not emphasized.

12. Sam and Jean are ice skaters for the school team, they have been practising hard for the state finals which are coming up next week. This is the most important competition of their life as the school has never found a championship before. With winning in their mind, they practised day and night with little rest, sometimes even skipping meals. Two days before the competition, the rival school appears and challenges them to a skate-off. Thinking that this would be a good opportunity to practise, they agreed. However, they did not realise that their skates were tampered with. As the duo were about to perform their trademark move, The Flying Swan. Jean's skates fell apart and she crashed to the ground, fracturing her leg.

HS.

Failure in the final sentence is not scored since it results from outside interference and does not reflect a lack of excellence.

13. The picture shows an ice skating rink where two people are in competition. Two sportswomen in action. The woman in the background is playing catch-up as she is closing up the leading woman. The woman in the lead appears confident of her chances of winning. They both want to win. The woman in the lead will lose her lead to overconfidence and complacency. They will receive their medals after the championship and celebrate together.

HS, HS, HS, HS.

"Lose her lead" is not scored for FF since it is later downplayed. The skater coming in second celebrates with the eventual winner, suggesting that the second place finish is not viewed as a failure.

14. A young man and lady sit on a wooden bench by the river bridge. The man stares at the girl with questioning eyes. The girl has told the man that she has to leave the city with her family and can no longer be with him. The man is filled with disbelief and asks, "Is there nothing I

can do?" The lady shakes her head and a tear escapes her eyelid. She stands up and starts to walk away but the man takes her hand. "Let's run! To a place far away from here!", exclaims the man.

None.

No achievement-related imagery.

15. It was a hot and sunny day at the court. Adam was against Steven in that match and it was the deciding shot for the match. Adam served a shot to Steven and as he attempted to return, he slid and tumbled over on the court. He missed the shot. Adam won the match. As Steven sat up, he held his hand to his head in defeat. He kept thinking, "Why did i trip?" The match ended and Steven lost. He walked off the court in defeat and shame.

FF, FF, HS, FF, FF, FF, FF

Falling in a sporting context counts as failure. Holding a hand to one's head in response to failure unambiguously indicates negative affect and so is scored for FF under the negative affect criterion. Asking the question "why did I trip?" is scored under the need to avoid failure criterion, since the player is wondering why he failed and how it could have been avoided.

16. Why is it so hard to do??!! The student felt, a sense of frustration and irritation for not being able to complete a simple exercise. Always an average student, she has yet to experience any major achievement in the academic sense. Feeling the pressure from others, and worst of all herself, each passing day is a stressful one... with much difficulty and help from friends, she finally be able to solve this question and move on to the other question.

FF, FF, FF

Both instances of negative affect are scored. Past lack of academic excellence viewed as failure by the character, and so is scored for FF. Success in final question not scored since it is a direct result of outside help.

17. The guy in the middle looks like he's in the middle of an exam. These people are probably the exam candidates currently in the middle of an exam. They are probably thinking about the exam questions, trying to recall the answers to the questions. The majority should be feeling an adrenaline rush or panic since that's the feeling exams give the most of us. However, some of them will definitely be feeling pretty confident so long as they have fully prepared for the examination. They all want to be able to pass with flying colours. They will probably continue writing their answers until time is up.

HS, HS

Panic is not scored for FF in this instance, since without further elaboration it neither implies expectation of failure nor clearly inhibits further achievement-related activity.

18. He is stressed out by his work. This is common in the 21st century workplace, as bosses frequently demand more than one's ability can supply. He looks very depressed and will probably burn out if he does not find an appropriate remedy for his crisis.

FF, FF.

Either the negative affect or the expected failure criteria can be scored in the third sentence.

19. These people are in their lecture theatre sitting in on a lecture after they had their lunch break. It's two in the afternoon and they are feeling full and happy after a good lunch. But once the lecture starts, the students scramble to keep up with their professor as he is explaining the topic to them. It is difficult for them to understand as the topic is very difficult and they are getting frustrated and stressed out. They are thinking that they will have to go back and revise to be able to keep up with the rest of the class. So, they rush home after school to study.

FF.

Difficulty in understanding is not strong enough to warrant scoring FF under the failure criterion here. As such only the negative affect is scored.

20. Four people are playing a basketball match. The player in the centre is clearly thinking about taking the shot, disregarding his teammate who is standing right behind him and potentially has no obstruction. His opponents on the other hand are so intent on blocking his move, perhaps knowing that his is the better player that they have ignored his teammate, leaving him potentially open to make his move. The situation is tense as one team is trying their utmost to prevent the other from scoring. However, despite appearing that he has ignored his partner, it was all part of his plan and his next move was to pass the ball back, feigning a shot and so taking his other two opponents out of the equation and allowing his team mate to score.

HS, FF, FF, HS.

Oftentimes descriptions of competitive sport will include achievement-related thoughts and actions relating to both competitors or teams. In such cases, all achievement-related imagery is scored from the perspective of the person or persons who are currently being described (e.g., "intent on blocking his move" is scored for FF under the need to avoid failure criterion since the opponents are the subject of the sentence.)

21. A bunch of people are fighting, and one of them has killed a guy. The culprit flees as fast as possible. The rest are shocked, not knowing immediately how to respond. In the end, all of them escape, and the poor man just dies like a fly.

None.

No achievement-related imagery.

22. Two males playing squash. Although it is impossible to know just by looking at the picture, it seems as if the guy in the background is anticipating the shot, while the guy in the foreground is concentrating on hitting the ball. Both will be thinking on how to counter each other.

None.

No clear standard of excellence.

23. Two persons are ice skating. They are competing against each other to see who is the fastest. They seem very happy while doing it.

HS.

Being the fastest is a clear achievement goal, and so the participants' pursuit of it is scored for HS. However the positive affect (enjoyment) seems to result from participation rather than progress or mastery, so HS is not scored under the positive affect criterion.

24. This is an examination hall. People are doing an exam. Everyone is concentrating on the questions given. Some of them are looking around. They might finish the exam paper in time.

None.

No clear standard of excellence.

25. The exam is in progress, and the students are working on it. Some of them are feeling nervous while others are confident. They will finish the exam.

FF, HS.

Both nervousness and confidence warrant scorings of HS and FF under the expectation of success and failure criteria.

26. Jeremy had stayed up all night to study for his final year exam today. Feeling sleepy and tired, he could not concentrate on his paper. He pondered hard over the question and tried to recall what he has studied last night. As the minutes pass, he failed to answer the question. Demoralised and upset, he browsed through the papers trying to find questions that he could answer. At the end of the paper, he shook his head and walked out of the exam hall slowly.

FF, FF, FF.

The shake of the head in the final sentence betrays either expected failure or negative affect, and so is scored for FF.

27. She is trying to do the assignment given by her teacher in the class. She will have to submit the project by the end of the lesson, and now she is searching for inspiration to write things down. The second she gets an idea, she writes down everything she can think about. After finishing the word quotas, she is planning to submit her work to the teacher and go out of the class earlier.

None.

No standard of excellence, just task imagery.

28. These are ice skaters. They are just concentrating hard and focusing on skating and doing their best. This might be at a competition or rehearsal of some sort. They hope to win if it is a competition. They also seem to feel tense and stress as their faces look serious. I believe that they will win.

HS, HS.

Although it is not definitively stated that the skaters are participating in a competition (rather than a rehearsal), these instances of motive imagery are still scored, in accordance with the general rules.

29. It is 7pm on the clock, and he sat at his desk, rubbing his temples tiredly. He thought to himself, "How will I ever finish this heap of work that keeps piling up?" The deadline for the tender is tomorrow and he is barely done with it. Work commitments aside, his wife and children are waiting for him to come home to celebrate his daughter's 3 year old birthday. But how could he leave with the amount of work awaiting him? Tensions were building up at home with his wife as he always returned home late from working over-time. He knew his wife would be unhappy if he did not leave the office for home very soon. The cell phone beside him started to ring. He looked at it for a moment, rejected the call and went back to his work.

FF.

Doubt as to whether the deadline can be met is scored for FF. Although the worker is clearly striving to meet external demands (i.e. the deadline), his sacrifice heavily suggests this goal is self-endorsed. However, if we were striving to meet the deadline in order to avoid his boss' wrath, this same sentence would not be scored.

30. It is yet another long night for Dave. He had been working overtime for the past two weeks, at a job he simply hated despite putting on all smiles for his colleagues and superiors all the time. His dream was to become an artist, not sit at a desk doing some stinking nine to five job which sucked the life out of him. He thought about his parents and how they had struggled to put him through his college education, his wife who has always been so proud of him for getting promoted to his current position so quickly, and his children who seemingly had unlimited wants and desires for the newest and most expensive toys. He was tired of it all, yet he could not just stop. He smiled faintly to himself as he thought about what he was going to do next. He had planned it for the past year, making sure to tie up his loose ends.

None.

No standard of excellence. The employee clearly does not view his current job as achievement-related and only works hard in order to satisfy external, affiliative demands (wife, children). His dream of becoming an artist could be scored for HS, but only if it were stated that he wished to excel as an artist, rather than simply change careers.

31. A surgery is going to start, or perhaps it's just finished. Or maybe it's a medical lab in a university. There is either a surgeon and his assistant, two surgeons, or a professor and a student. They want to communicate with each other. Next they will leave the room, start the surgery, or continue to talk

None.

No standard of excellence.

32. The group is having a mid-term examination in the lecture theatre. The theatre is quiet, with everyone fully focused on the examination. The main man in the picture appears to have an expression of worry. He has prepared well for the examination, but the questions do not appear to be familiar. He tries his best. He is thinking about guessing his way through the questions he is unfamiliar with. He can only do so much to better his chances of doing well in the examination. He finishes the paper early and leaves.

FF, HS.

Worry clearly relates to the achievement outcome here, and so is scored as FF under the expectation of failure criterion. The student "trying his best" describes his commitment to the achievement goal, and so is scored for HS under the need for success criterion. "He can only do so much to better his chances" is not scored under the failure criterion since the possibility of failure is only hinted at rather than heavily implied. If it was clear that the student did not know the answer to any of the questions then left early, the final sentence would be scored for FF.

33. This is a class whereby the people are waiting to complete a test. Final year paper is upon them and the atmosphere is tense and only scribbling can be heard in the class. All the students are intent on completing the paper and achieving the best grade as possible. Small Clicks can be heard from the constant walking of high heeled professors and the anxious movements of the students who are trying their best to recall the massive loads of information they have tried to memorize the night before. Meanwhile those that have studied sufficiently take their time and slowly complete the paper one page at a time. Before long, the bell signalling the end of the paper sounds and the whole hall erupts into a sigh of relief. It is finally OVER!

HS.

Scored under the need for success criterion.

34. A girl is doing a math exercise in class. The exercise is very easy for her and she can come up with the solution in a short time. However, the two boys behind her are discussing the questions because they cannot finish it them by themselves. They all finish the exercise in the end.

None.

No standard of excellence.

35. It's a normal day in the hospital with a couple of surgeries on going. Dr John, on the left, is a general surgeon performing a simple surgery for his patient. He cut open the patient, fixed the problem, and is now closing the patient up. His intern, Christina, watches over and assists him as much as she can. Dr John hopes that by watching and gradually helping out in the surgeries, Christina would learn more and become a great surgeon. Christina watches on and is amazed at Dr John's fine stitching skills. She aspires to be like him.

HS, HS.

The initial description of the surgery should not be scored for achievement imagery since it is

described as being routine. Aspiring to be like the great surgeon is scored for HS under the need for success criterion.

36. It was finally the final examination. John had studied very hard for it, and he had to do well. As he stared at the paper, he felt nervous and tried to calm himself down. The paper was difficult, just like everyone had expected. He looked around, and saw that his course mates were racking their brains doing the paper. He decided to focus on his own work, and started doing it. It wasn't as hard as he initially thought. He finished the paper and felt relieved.

HS, FF.

Feeling nervous is scored for FF here since it is linked to the appraisal of the paper as being difficult in the adjacent sentence. When nervousness arises from perceived difficulty, it is safe to assume a least partial expectation of failure, and therefore to score FF.

37. The picture shows a girl writing something with her pencil, probably taking a test. In the background there are two boys. The girl is concentrating on her work, and the boys are interacting with each other. The girl is calm and collected. The girl wants to complete her work and is not distracted by external stimuli and looks focused. The teacher will come along to collect her paper.

None.

No standard of excellence.

38. A female student taking notes in a class while the two guys in the background are busy talking. The female student will be thinking of understanding what the person she is listening too saying, while the two guys are thinking about their conversation instead.

None.

No standard of excellence.

39. The surgeon's hands shook lightly, barely visible to anyone, but he knew it himself. The man on the patient's bed was his father, the man who brought him up singlehandedly and gave him his education. Now he was lying here in pain, pleading with his eyes for his son to put him out of his misery. "What if they found out?" A million thoughts raced through the surgeon's mind, and he struggled to focus on the procedure he had rehearsed so many times. He could not afford to make any mistakes.

None.

No achievement-related imagery. The goal here (ending his father's suffering) is affiliative in nature.

40. This is a test happening in the lecture theatre. The guy in the centre of the picture is stumped - he does not know the answers to the foreign looking questions on his paper. He is feeling anxious and wishes that somehow, the answers would magically appear in his head. Unfortunately, that did not happen and he left the lecture theatre calculating how well he would have to score in the upcoming tests so that a B+ is still possible for this module.

FF, FF, FF.

Not knowing the answers to the questions is a partial failure and so is scored for FF. Feeling anxious clearly relates to the future outcome of the exam or the inability to answer the questions, so can be scored either under the expectation of failure or negative affect criteria. The second reference to his inability to answer the questions is also scored for FF under the failure criterion, although the final sentence could also be scored as a need to avoid failure, since the student hopes that future good performances will mitigate the effects of his poor showing in the current test.

41. This is an ice-skating race. The first two skaters are sprinting to the final. Both of them want to get the champion. The second skater wants to overrun the first one so as to win the race.

HS, HS.

Both sentences are scored under the need for success criterion.

42. Guy: "What? You are pregnant?"
Girl: "Yes, I just verified it today."
Guy: "Oh no, what are we going to do about it? Let me digest this first. You are pregnant, but we are still so young! We are still in school! I don't have the financial means to support a baby, let alone you!"

Girl: "I know, but..."
Guy: "And our parents can't know about this! My dad will kill me! No, wait, I think your mum will kill me first!"
Girl: "You are not making me feel better..."
Guy: "Listen, we have to solve this like adults. I will find a gynaecologist. You should get an abortion ok? We are still so young! Think of the restrictions of being a mother at such a young age! Listen to me ok?"

None.

No achievement-related imagery.

43. Two surgeons are preparing for huge operation. They have been doing operations for years and years but still each operation makes them feel nervous. They are carefully looking through all the devices that they need for operation and they strongly wish they can save another life! Time for the entry of patient. Good luck!

HS, HS.

Here nervousness is not clearly the result of expected or anticipated failure, and so is not scored for FF.

44. Mother and daughter take to the ice rink and show off their skills. They become like two identical figures on the ice. Sponsored by the best of companies, their success has reached a pinnacle and it seems like nothing can stop them. Anna is not happy though, as she longs to be with the love of her life. That evening she tells her mother she loves her boyfriend Hamish and will return to him in Norway. Her mother looks at her with contempt, for Hamish is Indian and therefore not a suitable match. Her mother's prejudices have haunted Anna all her life,

and she has had enough. She strangles her mother with a traditional Indian sari. It's a nice touch.

None.

No achievement-related imagery. Success in skating is downplayed, since it is superseded by affiliative (daughter being with boyfriend) and power (mother controlling daughter's life) goals which are the focus of this story.

45. There is always a need to improve and learn on the job. This is the motto that the Chief of Surgery Mercy West lives by day to day. Today, he is about to guide me, a first year resident, in how to carry out a proper endoscopy. I had not previously done this procedure and to my awe I had been chosen to assist him in the surgery. An hour before the surgery, he called me in to the operating theatre and asked me to prep myself for surgery. I did so, even though I found it peculiar that we were doing so pretty early. I then found out that he was going to do a dry run and a walk through of the surgery. It was through this that I got to visualise what he expected at each stage of the surgery. I was really impressed at how the Chief was so meticulous in his job and at the same time found time to teach me the proper way of doing things.

HS.

The need for improvement in the first sentence is later attributed to the chief of surgery, and so is scored for HS.

46. This is an exam scene in a lecture theatre, and it focuses on a single guy (let's call him Mike). Mike has been studying the entire night before for this quiz after choosing to party the week away, and right now in the midst of the quiz he struggles to recall what he has been studying for so frantically. He tries to piece together the information lying somewhere in his brain, and reproduce them as best as he can on the script. He hasn't had the time to shave, and his stubble shows. The questions do not make sense, the answers aren't forthcoming, and he knows right now that he should have started studying earlier. Regret does not help him however, and he resolves to study earlier next time. But for now, this quiz would most likely return a bad result, and there is only so much he can do to salvage it.

FF, FF, FF.

"Studying the entire night before" and "studying for so frantically" are not scored for HS since they are effortful activities performed in order to avoid a negative outcome (failing the exam). This can be safely assumed given that the student in question did not start studying until the night before the exam.

47. A dating scene by the riverside where a couple is sitting on the bench talking. There is a guy and a girl which suggest that they are probably dating. The guy is probably trying to impress the girl.

None.

No achievement-related imagery.

48. This is first ice skating competition organised by McDonald's, as we can see their logo on the competitor's head band. It is time for the girl's ice skating and there are only two competitors. McDonald's totally failed to make this competition successful. However, with their sportsmanlike spirit, these two women are trying their best not to be on the second place which is equals to last place. They both want to win, but Sally who is skating fastest got first place!

FF, FF, HS.

Since organizing a successful competition would most likely be a work-related achievement goal for the employees responsible, McDonald's failure to make the competition a success is scored for FF under the failure criterion. The two skaters "trying their best" is scored for FF as it indicates pursuit of a negatively-framed achievement goal.

49. They have been in surgery for 6 hours now. The surgeon had just transplanted the liver to the patient. All was going well. Just then, the patient began having seizures as his body jerked violently. The whole operation room was thrown into a frenzy as they tried to figure out a solution. The nurses tried to gain control over his condition as the surgeon continued the procedure. When it seemed that all hope was lost, the patient suddenly showed signs of calm and returned to a normal state.

HS, HS, FF, HS.

Partial success in the third sentence is scored for HS. FF is later scored for the expected failure, while saving the patient is scored as HS under the success criterion.

50. Amy was falling behind with her school work and her grades were dropping. Her parents were disappointed with her and told her to pull her grades up from now on. She felt irresponsible and decided to focus more in class and work harder. This picture shows her concentration and focus while doing her work in class. She knows she can do well, and she wants to do herself and her parents proud. She sits for the examination and manages to pull her grades up to become the best student in the class. Finally, her hard work has paid off.

FF, FF, HS, HS.

The initial failure is discounted since Amy eventually becomes the best student in the class. The parents' disappointment in response to her failure is however scored under the negative affect criterion. "Deciding to focus more in class and work harder" is not scored since it is a direct consequence of her being admonished by her parents. However subsequent imagery is scored since it is clear that this standard of excellence has been internalized.