



COPYRIGHT AND USE OF THIS THESIS

This thesis must be used in accordance with the provisions of the Copyright Act 1968.

Reproduction of material protected by copyright may be an infringement of copyright and copyright owners may be entitled to take legal action against persons who infringe their copyright.

Section 51 (2) of the Copyright Act permits an authorized officer of a university library or archives to provide a copy (by communication or otherwise) of an unpublished thesis kept in the library or archives, to a person who satisfies the authorized officer that he or she requires the reproduction for the purposes of research or study.

The Copyright Act grants the creator of a work a number of moral rights, specifically the right of attribution, the right against false attribution and the right of integrity.

You may infringe the author's moral rights if you:

- fail to acknowledge the author of this thesis if you quote sections from the work
- attribute this thesis to another author
- subject this thesis to derogatory treatment which may prejudice the author's reputation

For further information contact the University's Director of Copyright Services

sydney.edu.au/copyright

**The Capability to Align Estimated Performance with Actual
Performance: Insights from Physical & Cognitive Performance
Contexts**

Tate Nolan Hubka

B.Sc. Exercise Science

Submitted in fulfillment of the requirements for the degree of Masters of Applied Science

Discipline of Exercise and Sport Science

Faculty of Health Science

The University of Sydney, NSW, Australia

June 2015

Declaration of Originality/Authorship

I, Tate Hubka, hereby declare that the Masters by Research thesis entitled ‘The Capability to Align Estimated Performance with Actual Performance: Insights from Physical & Cognitive Performance Contexts’ is no more than 60,000 words in length including quotes and exclusive of tables, figures, appendices, references, and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. This thesis contains material that has been submitted for publication in peer-reviewed journals. This thesis is entirely my own work, and except where due acknowledgement is made, no other person’s work has been used.

Tate Hubka

A handwritten signature in black ink, appearing to be 'TH' with a stylized flourish.

Date: June 1st, 2015

Abstract

Discrepancies between estimated and actual performance occur daily in both normative and performance based tasks. This is synonymous with a type of cognitive bias known as the Dunning-Kruger Effect (DKE). In this thesis, Chapter 2 examined the existing literature on estimation and performance alignment and DKEs using systematic and meta-analytical procedures. Findings identified a small-moderate correlation in the ability to align estimation with actual performance. In DKE terms, quartile 1 performers overestimated, while quartile 4 underestimated. Alignment correlations were also found to be moderated by methodological and task factors, but not participant characteristics.

Chapter 3 assessed DKE prevalence and whether sporting experience, the time point of estimation, and reference group moderated trends in the physical tasks of Sprint and Vertical Jump. Notwithstanding DKE presence, trends were affected by time point of estimation. Estimation error was not related to current or previous sporting experience in either task.

Chapter 4 examined DKEs in the cognitive contexts of the Stroop and Tower of Hanoi tasks, and assessed whether estimation error was moderated by time point of estimation, reference group, task difficulty, feedback, and efficacy. For both tasks, pre-task efficacy predicted estimation error, and time point of estimation affected estimation, with increases and decreases post-task in the Stroop and Tower of Hanoi respectively.

Together, findings highlight DKE prevalence in multiple task contexts. DKE trends were moderated by task and methodological characteristics. Underlying mechanisms appear to implicate metacognitive skill as well as chronic-self views and pre-task efficacy. Increasing metacognitive skill and performance feedback is identified as a key strategy for error prevention and mitigation. Identifying DKE consequences and interventions that improve estimation-performance alignment are important future directions.

Keywords: Perceived Performance, Actual Performance, Dunning-Kruger Effect, Meta-Analysis, Self-Assessment, Metacognition

Preface

All of the work presented henceforth was conducted at the University of Sydney, Cumberland campus. The University of Sydney's Human Research Ethics Board approved all projects and associated methods.

Acknowledgments

This thesis is the culmination of my research as conducted from March 2013 to February 2015. Although this thesis details the research design, data collection, data analysis, and interpretation of my research, it by no means adequately summarizes the degree of effort, work, or assistance I received throughout my Masters degree, nor which went into this body of work. Therefore, I would like to specifically acknowledge those whose assistance has allowed me to reach this far.

I would like to primarily thank my supervisor: Steve Cobley, for his constant assistance, encouragement, and patience. Due to this, I was able to involve myself in teaching undergraduate courses, and additional research collaborations. These tasks increased my passion and enjoyment of academia, from which I intend to continue into a PhD. I would also like to acknowledge the assistance provided by my supervisors Roger Adams, who provided help with understanding the complexities of statistics, and Donna O'Connor, who provided assistance with research dissemination, as well as providing additional research experience.

Additionally, I would like to acknowledge (in no particular order), Cristian Alegria, Fiona Pearson, Vindy Jayewardene, Matt Hoon, Allyson Attenborough, Helen Parker, Cloe Cummins, Scott Michael, Alycia Fong Yan, Tim English, Nate Morris, Shelley Davison, Claire Tompsett, and Caitlin Taylor. They made my stay at the University of Sydney very enjoyable. Finally I would like to thank Kritika Rao, for her support and encouragement for as long as I have known her.

To all of you, and any I may have forgotten: Thank you, very, very, much.

Table of Contents

Declaration of Originality/Authorship.....	i
Abstract.....	ii
Preface	iii
Acknowledgments	iii
Table of Contents.....	iv
List of Tables	ix
List of Figures.....	x
Chapter 1 - Introduction.....	1
1.1 Research Focus	1
1.2 Research Background	3
1.3 Research Perspective	4
1.4 Research Aims.....	5
1.5 Research Significance.....	6
Chapter 2 – The Capability to Match Estimated Performance with Actual Performance: A Meta-Analysis.....	8
Abstract.....	8
2.0 Introduction	9
2.0.1 Research inconsistencies.....	10
2.0.2 Additional explanations	11
2.0.3 Meta-analysis purpose	12
2.1 Methods	13
2.1.1 Systematic search.....	13
2.1.2 Inclusion criteria	13
2.1.3 Data extraction.....	14
2.1.4 Data analysis	16
2.2 Results	19
2.2.1 Systematic search.....	19
2.2.2 Meta-analysis	26
2.2.2.1 Overall.....	26
2.2.2.2 Correlation moderator analyses.....	30

2.2.2.3 Mean difference moderator analyses	31
2.3 Discussion.....	32
2.3.1 Meta-analysis findings	32
2.3.2. Findings from systematic reviewing.....	36
2.3.2.1 Psycho-social characteristics.....	36
2.3.2.2 Consequences.....	37
2.3.2.3 Interventions.....	38
2.3.3 Recommendations and future directions.....	39
2.3.4 Limitations	39
Chapter 2 - 3 Bridging Statement	41
Chapter 3 – The Capability to Align Estimated and Actual Performance: The Dunning-Kruger Effect is Evident in Physical Task Contexts	43
Abstract.....	43
3.0 Introduction	44
3.1 Methods	46
3.1.1 Participants.....	46
3.1.2 Procedure	46
3.1.2.1 60m Sprint.....	47
3.1.2.2 Vertical Jump	47
3.1.2.3 Task estimation	47
3.1.3 Data analysis	48
3.1.3.1 Overall estimation alignment	48
3.1.3.2 Estimation-performance difference.....	50
3.1.3.3 Estimation-performance error	50
3.1.3.4 Estimation moderation	50
3.1.3.5 Task experience.....	51
3.2 Results	51
3.2.1 60m Sprint.....	51
3.2.1.1 Overall estimation alignment	51
3.2.1.2 Estimation-performance difference.....	52
3.2.1.3 Estimation-performance error	52
3.2.1.4 Estimation moderation	52

3.2.1.5 Task experience.....	57
3.2.2 Vertical Jump.....	57
3.2.2.1 Overall estimation alignment.....	57
3.2.2.2 Estimation-performance difference.....	58
3.2.2.3 Estimation-performance error.....	58
3.2.2.4 Estimation moderation.....	58
3.2.2.5 Task experience.....	60
3.3 Discussion.....	60
Chapter 3 - 4 Bridging Statement	65
Chapter 4 – The Dunning-Kruger Effect in Cognitive Performance Contexts	67
Abstract.....	67
4.0 Introduction.....	68
4.1 Methods.....	72
4.1.1 Participants.....	72
4.1.2 Procedure.....	72
4.1.3 Data analysis.....	75
4.1.3.1 Overall estimation alignment.....	75
4.1.3.2 Estimation-performance error.....	75
4.1.3.3 Estimation moderation.....	76
4.1.3.4 Task efficacy.....	77
4.2 Results.....	77
4.2.1 Stroop.....	77
4.2.1.1 Overall estimation alignment.....	77
4.2.1.2 Estimation-performance error.....	80
4.2.1.3 Estimation moderation.....	80
4.2.1.4 Task efficacy.....	82
4.2.2 Tower of Hanoi.....	82
4.2.2.1 Overall estimation alignment.....	82
4.2.2.2 Estimation-performance error.....	85
4.2.2.3 Estimation moderation.....	85
4.2.2.4 Task efficacy.....	86
4.3 Discussion.....	88

4.3.1 Stroop.....	88
4.3.2 Tower of Hanoi.....	90
4.3.3 Implications and recommendations	92
Chapter 5 – A Discussion of Dunning-Kruger Effect Studies in Physical and Cognitive Contexts	94
Abstract.....	94
5.0 Introduction	95
5.1 Study Summary	95
5.2 Analysis of findings.....	97
5.2.1 Egocentrism	98
5.2.2 Task efficacy.....	99
5.2.3 Chronic self-views	100
5.2.4 Task specific skill	101
5.2.5 Metacognition	102
5.3 Statement of Originality	103
5.4 Implications	104
5.5 Limitations.....	106
5.6 Future Directions	107
Chapter 6 - Conclusion.....	109
References.....	111
Appendices	126
Appendix A – PRISMA Checklist.....	126
Appendix B – Example list of excluded studies.....	128
Appendix C – Human Research Ethics Committee Approval	130
Appendix D – Study Advertisement.....	132
Appendix E – Participant Information Statement/Consent Form.....	133
Appendix F – Pre-Study Questionnaire.....	138
Appendix G – Pre and Post Task Questionnaires.....	139
Appendix H – Data Collection Sheet	145
Appendix I – Participant Debrief.....	146

Appendix J – Summary of Masters Candidature Contribution 147

List of Tables

Table 1: Illustrates studies identified from the systematic review process, presented chronologically and according to participant characteristics, task domain, performance test, DKE identification, and explanation(s) of findings.....	20
Table 2: Equations of combined Estimation-Time and Reference-Group variables.....	49
Table 3: Actual and combined estimated percentile ranks for 60m Sprint and Vertical Jump according to performance quartile.....	53
Table 4: Overall and quartile specific correlation coefficients between Actual Performance Rank and each combined estimated performance rank for the 60 m Sprint and Vertical Jump.	55
Table 5: Regression slopes for actual and combined estimated performance percentile ranks for 60m Sprint and Vertical Jump according to Performance Quartile.....	56
Table 6: Comparison of estimated performance ranks types (pre v post; student v athletes) in Sprint and Vertical Jump according to performance quartile.....	57
Table 7: Actual and estimated percentile ranks for Stroop Test and Tower of Hanoi according to performance quartile.....	78
Table 8: Comparison of estimated performance ranks types (pre v post; student v athletes; efficacy) in Stroop Test and Tower of Hanoi according to performance quartile. ...	81
Table 9: Task specific results and characteristics as per Chapter 3 and 4 findings.....	98

List of Figures

Figure 1: Systematic review steps & flow diagram adapted from PRISMA guidelines (Moher et al., 2009).....	15
Figure 2: Summary chart overviewing participant demographic, psycho-social characteristics, methodological approaches, actual performance tasks, individual consequences, & mechanisms associated with the DKE.	25
Figure 3: Forest plot of correlations between estimated and actual relative performance across and within studies.	27
Figure 4: Forest plot of quartile one (Q1) performer’s overestimation across and within studies.	28
Figure 5: Forest plot of quartile four performer’s (Q4) underestimation across and within studies.	29
Figure 6: Actual and estimated performance ranks for Sprint performance according to Performance Quartile.	54
Figure 7: Actual and estimated performance ranks for Vertical Jump performance according to Performance Quartile.	59
Figure 8: Actual and estimated performance ranks for Stroop performance according to performance quartile.....	79
Figure 9: Linear regression between Pre and Post-Performance Efficacy and Estimation Error in Stroop task.	83
Figure 10: Actual and estimated performance ranks for Tower of Hanoi performance according to performance quartile.....	84
Figure 11: Linear regression between Pre and Post-Performance Efficacy and Estimation Error in the Tower of Hanoi.	87

Chapter 1 - Introduction

1.1 Research Focus

Whether attempting to cross the street, safely drive a car, operate machinery or mechanical tools, or acquire new skills and knowledge in educational, sporting, or workplace settings; the ability to accurately perceive, comprehend, understand, and estimate actual performance is of vital importance (Eva, Cunnington, Reiter, Keane, & Norman, 2004). This ability is needed on a daily basis, influencing our capability regarding information utilization, which allows for situational awareness (e.g., looking both ways before crossing the street; Ehrlinger & Dunning, 2003); task analysis, which allows for determination of the safety of a situation (e.g., determining if the street is dangerous); personal skill assessment, which determines task readiness (e.g., knowing whether you are able to cross the street); decision-making, (e.g., deciding whether to cross the street; Mishra, 2014); strategy formulation, leading to the appropriate use of tools (e.g., taking advantage of traffic lights); and behaviour planning, related to the investment into specific tasks (e.g., planning how to cross the street). Together, these cognitive processes allow for a seamless transition from the desire to cross the street, to being on the other side successfully, a task in which most people do question or consider their ability to safely perform.

In contrast to somewhat more popular perceptions however, accurately aligning self-estimations with actual performance in other task settings is challenging and difficult; and since the 1940s, research studies have sporadically examined the relationship between estimated and actual performance. For instance, motivated to determine whether students could accurately evaluate their strengths and weaknesses, Arsenian (1942) asked USA college freshmen to rate their scholastic aptitude/achievement on a six point scale, ranging from the lowest 10% (very inferior) to the highest 10% (very superior), relative to all other

freshmen across the USA. Freshmen then completed the American Counsel of Education (ACE) psychological examination as a measure of aptitude/achievement. Based on perceived ability, freshmen were separated into quartiles and compared with their actual exam performance. Findings identified that self-estimates did not correspond well to exam performance (contingency coefficients = 0.26 – 0.57). Arsenian also illustrated associated (social) consequences for individuals who were most inaccurate aligning estimations with performance (i.e., scoring on average 9 points lower on the ACE exam). These included individuals being more likely to be dismissed or advised not to return at the end of the year; be in their second attempt at college; be categorized as maladjusted (i.e., having a “personality problem”); and be on a form of medical prescription.

Since 1942, meta-analyses have quantitatively identified small-moderate pooled correlations between estimated ability (task context & general) and actual performance (i.e., $r = .34$, Freund & Kasten, 2012; $r = .29$, Mabe & West, 1982). This suggests that the inaccurate alignment between perceptions and reality consistently occur, potentially leading to significant implications for those involved, some of which may be quite detrimental. For example, these implications can include, but are not limited to, motor vehicle and workplace accidents (De Craen, Twisk, Hagenzieker, Elffers, & Brookhuis, 2011), personal injury (Burson, Larrick, & Klayman, 2006), financial loss (Ferraro, 2010), and decreased motivation (Duda & Nicholls, 1992). This may then lead to a reduced knowledge and skill acquisition (Austin & Gregory, 2007; Gross & Latham, 2007), and impact on social interactions and relationships (Ames & Kammrath, 2004). This mal-alignment between estimations and actual performance is the focus of this thesis.

1.2 Research Background

Social and cognitive psychologists have long been interested in the topic of human error, and the capability to self-assess and evaluate performance. The misalignment of which can be considered to be an error in cognitive functioning. Such errors in perception have been found to exist in a diverse range of task and assessment situations, including perceived teaching ability in university professors (Cross, 1977), judgment of ability in professional soccer players (Vanyperen, 1992), diabetes knowledge in nurses (Baxley, Brown, Pokorny, & Swanson, 1997), factual knowledge in students (Eva et al., 2004), creativity in children (Kaufman, Evans, & Baer, 2010), driving ability of police (Waylen, Horswill, Alexander, & McKenna, 2004), as well as the driving ability of novice and experienced drivers (De Craen et al., 2011). This error can, for example, lead to over or underestimation of performance capability.

Traditionally, research in this area has identified and suggested that forms of cognitive bias can account for errors between perception and reality (Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008; Kruger & Dunning, 1999), especially when one's subjective estimation or assessment in a given task is consistently dissonant to more objective or quantifiable information (Kahneman & Tversky, 1972). As the cognitive processes behind error vary, with certain individuals or groups more prone to error, there exist multiple types of cognitive bias. For instance the tendency to interpret information in a way consistent with one's current beliefs is known as the confirmation bias (Mahoney, 1977); the belief that previous success increases the chance for future success is known as the hot-hand fallacy (Gilovich, Vallone, & Tversky, 1985); and the tendency to believe or do things because many others believe or do them (e.g., peer pressure), is considered the bandwagon effect (Nadeau, Cloutier, & Guay, 1993).

Cognitive biases can also be considered pervasive and chronic if they are repeated or exist across multiple situations, and relate to similar reasoning or processes (Ferraro, 2010; Kruger & Dunning, 1999). Chronic biases can manifest beyond processes of inaccurate judgement due to forms of perceptual distortion, illogical interpretation, and irrationality (Cohen, 1981); and so can also be associated with a compromise in cognitive functioning, or psychological disorder (Jahoda, 1958; Taylor & Brown, 1988). Further, the existence of biases indicate that discrepancies between estimated and actual performance do not necessarily provide a form of cognitive alarm, stimulating the development of knowledge or skill for error rectification; instead leading to a continued belief in the accuracy of one's performance beliefs, even when misconceptions and inaccuracies may be harmful (e.g., personal injury).

The underlying processes accounting for cognitive biases have historically been associated with: circumventions or short-cuts in information processing (i.e., heuristics - Kahneman, Slovic, & Tversky, 1982); motivational and social processes (i.e., attribution; Abramson, Seligman, & Teasdale, 1978; Heider, 2013); self-serving biases (e.g., hindsight bias - Arkes, Faust, Guilmette, & Hart, 1988); as well as limitations in knowledge or information processing capacity (e.g., Dunning-Kruger effect - Kruger & Dunning, 1999; false-consensus effect - Ross, Greene, & House, 1977).

1.3 Research Perspective

Cognitive bias studies have generally been explored using methodology and research designs common in social and cognitive psychology. For example, in order to determine cognitive errors specific to task performance, studies have acquired actual performance through the use of specific performance tasks (e.g., academic tests, questionnaires), while perceived performance has been acquired by asking participants to estimate their

performance. The dissonance between estimations and actual performance is considered the estimation error, and may be due to a bias known as the Dunning-Kruger Effect (DKE; Dunning, 2011; Schlösser, Dunning, Johnson, & Kruger, 2013).

Kruger and Dunning (1999) first found evidence for the DKE after independently evaluating undergraduate student competency in various contexts (i.e., humour, logic, grammar) using quantitative questionnaires, and obtaining students estimations of their performance in relation to the 'average class student' on a percentile rank (i.e., 0-100%). Kruger and Dunning (1999) found that those who performed in the highest quartile (i.e., top 25%, Q4), underestimated their performance by 11-19 percentile points; while those who performed in the lowest quartile (i.e., bottom 25%, Q1), overestimated their performance by 40-51 percentile points. This study uniquely highlighted the asymmetric estimation error associated with specific levels of actual performance, whereas previous research focused on the overall estimation-performance alignment. As such, this study suggests that not only are specific cognitive biases, and their associated mechanisms and moderators at work dependent on individual performance; but also the expression of these biases may differ according to performance quartile. As performance is an inherently malleable characteristic, the effect on individuals by the DKE can therefore be rectified.

1.4 Research Aims

The purpose of this thesis was primarily to examine overall individual self-assessment accuracy via the conduction of a systematic review and meta-analysis. This would allow for a descriptive and analytical discussion of what is (un)known regarding the DKE and its mechanisms, processes, moderators, limitations, and mitigating circumstances. The results of these informed the design of two additional studies to further investigate the DKE in unique contexts; while addressing additional individual, methodological, and moderator variables

suggested to influence self-assessment accuracy. Together, these studies attempted to determine the cognitive processes ongoing throughout self-assessment, while identifying its originality, implications, limitations, and proposed future research directions.

1.5 Research Significance

As previously mentioned, the mis-estimation made by individuals can affect daily decision-making, leading to drastic and severe consequences. Not only can these inaccuracies impact individual achievement (Mattern, Burrus, & Shaw, 2010), they can also lead to severe injury and death through overconfidence of performance capability (Burson et al., 2006; Palmer, 2002; Petrass, Blitvich, McElroy, Harvey, & Moran, 2012). Further, these consequences do not solely affect the individual making the error, but can cause similar implications for others, such as errors leading to deaths in surgical patients (Whitaker, 2008).

However, while considerable evidence exists to detail the inability to accurately self-assess performance capability, this evidence is varied in both the magnitude of estimation error, and the methodological design determining estimation error, making general recommendations seeking to rectify estimation error difficult. Further, due to the consequences of inaccurate self-assessments, an in-depth analysis regarding the cognitive bias known as the DKE is necessary, if understanding of the mechanisms behind inaccurate assessments is to be achieved. The contribution of such information to the current literature may also allow for the creation of effective mitigation strategies and intervention programs designed to increase individual estimation accuracy.

Further understanding of the cognitive processes ongoing throughout self-assessment, as well as the individual, methodological, and task specific moderators that mitigate estimation inaccuracy, may allow for specific recommendations regarding individual behaviour in order to decrease estimation error and associated negative consequences.

Additionally, recommendations for professionals such as teachers and workers may further build the awareness of mitigation strategies in students and employees, improving student learning, workplace efficiency, and individual safety.

Chapter 2 – The Capability to Match Estimated Performance with Actual Performance:

A Meta-Analysis

Abstract

Background: Error in self-assessment due to cognitive bias, specifically the Dunning-Kruger Effect (DKE) can lead to implications throughout one's daily life. **Objectives:** This meta-analysis was aimed to determine the capability to align estimated with actual performance, the pervasiveness of the DKE, and the influence of moderating variables. **Methods:** A systematic search of the literature combined with a meta-analysis was conducted investigating correlations and mean differences of individual assessment accuracy. **Results:** Results identified a small-moderate pooled correlation between estimated and actual performance ($r = .32$, $CI = 0.29-0.35$, $p < .001$), along with consistent pooled mean differences in DKEs. Specifically, Q1 performers consistently overestimated performance relative to actual performance (by 37.44 percentile points; $g = 2.17$, $CI = 1.74-2.60$, $p < .001$), while Q4's underestimated (by -19.96, $g = -1.22$, $CI = -1.43- -1.01$, $p < .001$), showing how mal-alignment is associated with performance competency. Pooled correlations were moderated by methodological (e.g., timing of estimation) and task (e.g., physical), but not participant related factors. **Conclusions:** Findings support Kruger & Dunning's (1999) metacognitive skill and false consensus explanations for estimation-performance inaccuracies. Consequences of mal-alignment and DKEs, and interventions that prevent or mitigate them are recommended as valuable future directions.

Keywords: Perceived Performance, Actual Performance, Dunning-Kruger Effect, Meta-

Analysis

2.0 Introduction

In their landmark study, Kruger and Dunning (1999) evaluated undergraduate psychology student competency in the areas of humour (by rating a list of jokes), logical reasoning (i.e., performance on Law School Admissions Test [LSAT] preparation guide questions), and grammar (i.e., performance on National Teachers Examination [NTE] preparation guide questions). After completing respective tasks, students estimated their performance relative to the 'average class student' on a percentile rank (i.e., 0-100%), and were grouped into actual performance competency quartiles (i.e., Quartile 1 = lowest 25%; Quartile 4 = highest 25%). Actual percentile ranks were then compared to self-estimated percentile ranks in each quartile and task. Across tasks, findings showed that the lowest performing quartile (Q1) significantly overestimated their performance by 40-51 percentile points, while high performers (Q4) underestimated their performance by 11-19 percentile points. This pattern of Q1 overestimation and Q4 underestimation subsequently became known as the Dunning-Kruger Effect (DKE; Dunning, 2011; Schlösser et al., 2013). The effect has been shown elsewhere, for instance quartile dependent error was evident in university students' logical reasoning ability (Ehrlinger et al., 2008), specialist physicians' clinical practice (Violato & Lockyer, 2006), and salesmen's ability to sell (Jaramillo, Carrillat, & Locander, 2003). Additionally, the DKE could account for the low moderate estimation-performance correlations, with Q1's and Q4's more likely to show error in their estimation relative to performance compared to Q2's and Q3's.

Explanations for DKE's differ according to performance quartile (i.e., Q1 v Q4). Based on their studies, Kruger and Dunning (1999) argued that Q1 performers suffered from a 'dual-curse' resulting from a combination of low task-specific and metacognitive capability. That is, limited domain specific knowledge and/or corruptions (e.g., overconfidence), led them to make more mistakes and errors during performance. Limited

metacognitive skill then renders them unable to recognize their own errors, the errors of others, or the fact that others may be making more appropriate decisions and/or may be performing better. In other words, Q1's inability to recognize their limited knowledge and skill leads them to believe that their performance is higher or at least comparable with others. By contrast, Q4 underestimation was attributed to 'the false-consensus effect' (Ross et al., 1977), and the belief held by Q4's that their peers are equally task-experienced and competent, leading Q4's to underestimate their performance (Kruger & Dunning, 1999; see Study 3 and 4).

2.0.1 Research inconsistencies

While correlations and DKEs have been identified, inconsistencies between studies have led to questions regarding the validity of findings, and raised debate as to the underlying mechanisms responsible. For example, estimation-performance correlations have ranged from small (e.g., $r = .14$; Sheldon, Dunning, & Ames, 2014) to large (e.g., $r = .55$; Ferraro, 2010). The degree of DKE mis-estimation by Q1 and Q4 performers has also varied considerably, from Q1 overestimation of 12 to 50 percentile points (Burson et al., 2006; Kruger & Dunning, 1999), to Q4 underestimations of -57 to -9 percentile points (Burson et al., 2006; Ryvkin, Krajč, & Ortmann, 2012). It is possible that such inconsistencies and variability may relate to multiple, but systematic, participant, methodological, and task-related factors between and within studies.

In terms of participant characteristics, inconsistencies may relate to age (e.g., mean age 41 yrs old: De Craen et al., 2011; v 24 yrs: Furnham, Kidwai, & Thomas, 2001), and gender (e.g., female: Battistelli, Cadamuro, Farneti, & Versari, 2009; v mixed: von Stumm, 2013). Influential methodological factors could be associated with variations in the estimation scale used in studies (e.g., likert scale: Mattern et al., 2010; v percentile rank: Sheldon et al., 2014), the type of performance evaluation applied (e.g., objective: Battistelli et

al., 2009; v subjective: Waylen et al., 2004), and how performance was categorized (e.g., tertiles: Gross & Latham, 2012; v quartiles: Kruger & Dunning, 1999). Finally, task type or context examined may also be influential, such as the difference between cognitive (e.g., Albanese et al., 2006; Ehrlinger & Dunning, 2003), and physical tasks (e.g., Mikkelsen, Kaprio, Kautiainen, Kujala, & Nupponen, 2005).

2.0.2 Additional explanations

While metacognition is viewed as the primary mechanism behind the DKE, additional mechanisms have been theorized to explain the differences in estimation accuracy due to participant, methodological, and task differences. Ehrlinger and Dunning (2003) suggested that to generate an accurate estimate of capability for an upcoming task, an assimilation of prior or related performance is hypothetically accumulated, leading to a stable self-view of capability (i.e., chronic self-views). In pre-task situations these chronic self-views may even be more indicative of actual performance (Ehrlinger & Dunning, 2003). When estimating performance immediately post-task however, chronic self-views may be influenced by immediate performance completion feedback. Rather than being produced by feedback from task components or adopted procedures within the task as would be expected (Marcora, 2009), post-performance estimates have been suggested to be due to the degree of effort invested (Taras, 2001), positive affect (i.e., feeling; Greifeneder, Bless, & Pham, 2010), and associated confidence (Felson, 1981) from simple task completion and/or a perceived successful performance attempt (Butler, 1990; Elzubeir & Rizk, 2000). Further, studies have suggested that task specific qualities such as task difficulty (Burson et al., 2006), available feedback (Ryvkin et al., 2012), task familiarity (Fitzgerald, White, & Gruppen, 2003; Mullen et al., 1985), and performance domain (Kruger & Dunning, 1999), may also influence the self-assessment accuracy of individuals.

While prior meta-analyses have provided pooled correlation estimates between estimated and actual ability (e.g., Freund & Kasten, 2012; Mabe & West, 1982), the sources of study inconsistency and potential moderating factors have not been identified. Further, no prior meta-analyses on DKE related studies have been conducted. Therefore, determination of systematic moderating factors, and the degree to which they affect both estimation-performance calibrations and DKE trends would be valuable. Delineating such influential factors (i.e., participant, methodological, and task), would provide a clearer understanding as to whom and when capabilities to align estimation with performance are affected. This would also help support or refute underlying explanations, explain how inaccuracies lead to detrimental consequences, and provide insight into possible prevention and intervention strategies that reduce estimation-performance mal-alignments.

2.0.3 Meta-analysis purpose

Using systematic search procedures (i.e., PRISMA guidelines; Moher, Liberati, Tetzlaff, & Altman, 2009) to identify studies, this meta-analysis was conducted to primarily determine a quantifiable pooled (across study) correlation estimate (i.e., overall estimation accuracy); and in terms of DKEs, determine an overall pooled mean estimated-actual performance difference estimate for both Q1 and Q4 performers. Such an analysis was carried out to identify the prevalence, consistency, and size of overall estimated-performance inaccuracies and the DKE. This would also allow for integration of correlation and mean difference methodologies. The second purpose was to determine whom, what situations/contexts (i.e., tasks), and what study methodological factors were more likely to be associated with consistent low/high correlations and DKE over/under estimations (i.e., moderating factors). Additionally, as part of data extraction and for purposes of summarising relevant and available studies, study features and data were tabulated and visually presented. This helped to identify consequences of low-correlation and DKE's; identify interventions

targeting the removal of self-assessment error; provide support to underlying mechanisms; and to establish foundations for future research.

2.1 Methods

2.1.1 Systematic search

A systematic literature search was conducted using the databases: CINAHL, EMBASE, Medline, PsychInfo, Scopus, and SportDiscus. These databases were utilized due to their relevance to psychological contexts, as well as the purpose of this study. Google Scholar was not utilized due to concerns regarding its reliability (Jacsó, 2005, 2008). Key words - in alphabetical order - included: better-than-average effect, or DKE, Dunning, Dunning-Kruger, or Dunning-Kruger effect, Kruger, metacognition, meta-ignorance, miscalibration, overconfidence, overestimate, perceived ability, perceived competence, self-estimate, unaware, underconfidence, underestimate, unskilled, unskilled unaware, unskilled-unaware effect, and worse-than-average effect. In order to evaluate self-assessment accuracy and the DKE in all eligible contexts, keywords were selected from known literature in order to allow for a broad search, and increased confidence in obtaining all relevant studies.

2.1.2 Inclusion criteria

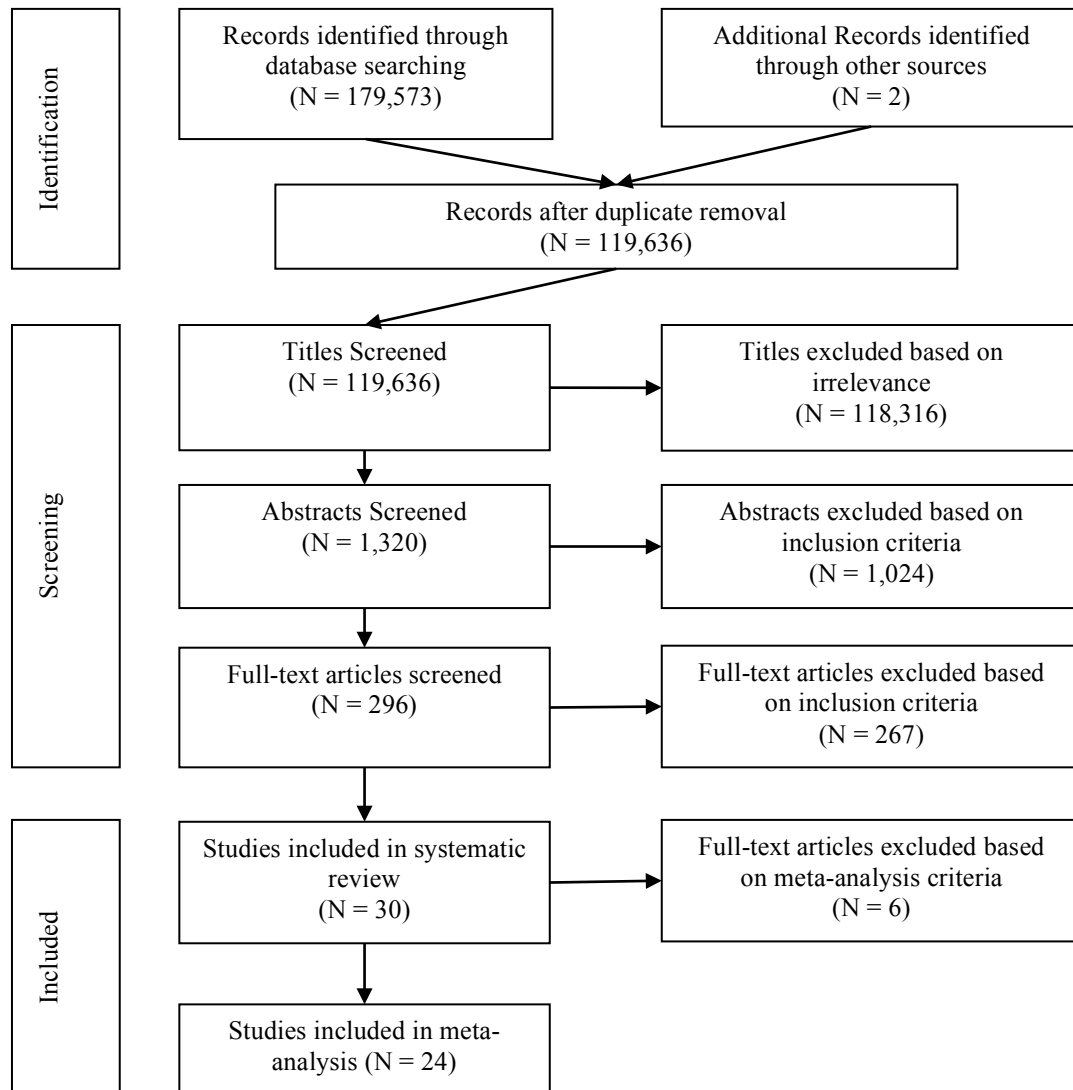
To be eligible for inclusion, studies had to primarily invite participants to estimate their specific task performance relative to others, and compare participant self-estimates to actual performance in the same task. Studies also had to contain a healthy adult population (18+ years), be published in peer-reviewed journals post 1940, and published in English. The systematic search process was adapted from the PRISMA statement and is shown in Figure 1. Unpublished non-peer reviewed articles (i.e., ‘grey literature’) were excluded, due to the possible additional bias produced from sampling various random studies with less than ideal methodological rigor (Baumeister, DeWall, & Vohs, 2009; Ferguson, 2010).

At each stage of the systematic search, studies were evaluated to determine whether they met inclusion criteria, and studies were rejected if they did not meet criteria. For example, studies were excluded if: they did not assess human performance, they did not assess estimated performance, or they used underage participants. If it was unclear whether a study should be rejected, the study continued through the PRISMA review stages until evidence for meeting or not meeting inclusion criteria became clear. Studies that met all review criteria and included quantifiable data were considered for statistical analysis, while those without quantifiable data were still descriptively analysed and reviewed. An additional team member, as a reliability check, reviewed search procedures. Examples of studies rejected can be found in Appendix B.

2.1.3 Data extraction

From the identified studies, data was extracted regarding: quantitative results, sample characteristics, influential psycho-social characteristics (e.g., narcissism), methodological approach, performance task or context, applied experimental manipulation (e.g., feedback), mechanisms (e.g., false-consensus effect), and potential consequences. Quantitative results included the overall correlation between estimated and actual performance (i.e., Pearson's correlation coefficients, r), mean differences between estimated and actual performance for Q1 and Q4 performers, mean difference standard deviations (SD), and accompanying t values. For sample characteristics, sample size, number of performance groups (e.g., quartiles), gender, age, experience, and status (e.g., university student) were extracted. For methodological approach, information was extracted which related to the: timing of self-estimates (i.e., pre v post performance); type of estimation scale (e.g., Likert scale v percentile); methodological consistency (i.e., consistent v inconsistent between estimated and

Figure 1: Systematic review steps & flow diagram adapted from PRISMA guidelines (Moher et al., 2009).



actual performance comparison); the nominated group against whom the estimates of performance were referenced (i.e., peers v average of a given task population); as well as actual performance comparisons (i.e., peers v average of population). Related to the performance task, both information related to the context (e.g., driving - De Craen et al., 2011; intelligence - Merkle & Weber, 2011), and measurement instruments or test applied (e.g., IPT-15 - Ames & Kammrath, 2004; LSAT - Kruger & Dunning, 1999) were recorded.

2.1.4 Data analysis

All extracted data was entered and analyzed using Comprehensive Meta-Analysis (CMA; Biostat, 2005). Study data, sometimes including numerous independent samples, was coded according to author, participant and sample characteristics, methodological approach, and task examined. If data could not be extracted, and/or not clearly or partially missing, then authors were contacted. A total of 17 authors were contacted for information, with 8 responding to provide appropriate and supplementary data. If authors did not respond, only available and relevant information within articles was used.

To determine an overall correlation estimate between estimated and actual performance, all relevant studies including background study (author) information, total sample size, and their reported correlation(s) were entered into a pooled correlation coefficient analysis (Borenstein, Hedges, Higgins, & Rothstein, 2009). Using Cohen's (1988) criteria, effect size of r was categorized as small (0.1), moderate (0.3), or large (0.5), while 95% confidence intervals (CI), along with respective p and z values were generated in CMA.

To identify DKE's overall consistency and prevalence, a second analysis was conducted using specific quartile groups. Estimated and actual percentile means and standard deviations (SD), and size of each quartile were entered. The raw unstandardized mean difference pooled point estimate (Bond Jr, Wiitala, & Richard, 2003; Borenstein et al., 2009), along with confidence intervals (CI), effect sizes (Hedge's g), z values, and p values were calculated for both the bottom (Q1) and top (Q4) quartiles of performance. Adapted from Cohen's d , Hedge's g effect sizes were categorized as small (0.2), moderate (0.5), and large (0.8).

To check for heterogeneity of results between studies, the I^2 and X^2 test of heterogeneity ($p < .10$) and visual inspections of the forest plots were used. I^2 describes the proportion of total variation in study estimates due to heterogeneity (Higgins & Thompson,

2002). An I^2 of less than 30% was used to indicate mild heterogeneity, and the use of a fixed-effect model; while an I^2 of more than 50% indicated notable heterogeneity, determining the use of a random-effects model (DerSimonian & Laird, 1986). Heterogeneity was identifiable in the correlation ($X^2 = 876.85$, $df = 40$, $I^2 = 95.44$, $p < .001$), and mean percentile difference analyses (Q1: $X^2 = 124.25$, $df = 21$, $I^2 = 83.10$, $p < .001$; Q4: $X^2 = 66.32$, $df = 21$, $I^2 = 68.34$, $p < .001$). Thus, a random-effects model was used in both analyses to combine data from independent studies.

Across analyses, publication bias was assessed using multiple tests including the fail-safe N (Rosenthal, 1979), and the 'Trim and Fill' method (Duval & Tweedie, 2000), along with visual inspection of the funnel plots. Fail-safe N determines the number of samples that would need to exist - which contained an average null result - for the main results to be due to sampling bias. Fail-safe N's for the correlation analysis were 8212, as well as 7065 and 3500 for the Q1 and Q4 mean differences analysis respectively. With the 'Trim and Fill' method failing to trim any studies, this suggests that studies were symmetrically located around the funnel plot. Collectively, along with funnel plot inspection, these tests suggest no evidence of publication bias.

Due to substantial heterogeneity in the pooled correlation and mean difference analyses, suggesting that differences between studies and their sample populations were related to systematic factors, sources of heterogeneity were investigated as recommended (Sutton, Abrams, Jones, Sheldon, & Song, 2000). As most variables were discrete, binary sub-categories were created and compared to identify significant moderating variables due to their simplicity and interpretability (Cooper, Hedges, & Valentine, 2009). For all sub-categories, a minimum of five samples was required to generate a pooled estimate. For example, in assessing the influence of gender on overall correlation estimates, one study reported six independent samples in six different tasks, permitting an analysis.

Binary comparisons for moderation analyses included participant characteristics of: gender (i.e., male v female), age (i.e., participants ≥ 40 yrs v < 40 yrs), and status (i.e., student v non-student). Due to data limitations, moderation effects of participant experience and psycho-social characteristics could not be investigated. However, age and status were used as proxies for experience, with older non-students considered as having more experience than young students. For methodological moderators, comparisons included: timing of self-estimation (i.e., pre v post actual task), the type of estimation scale used (i.e., Likert v percentile; Likert v bell curve; percentile v bell curve), and methodological consistency (i.e., consistent v inconsistent estimated & actual performance comparison groups). Different combinations of methodological (in)consistency were also examined, including estimated performance reference (i.e., peers v average), and actual performance comparison (i.e., peers v average). Finally, task performance type (i.e., physical v cognitive, physical v academic, academic v cognitive), was examined. With the exception of correlation categories examining gender, status and methodological variation (i.e., using a bell curve scale) heterogeneity was evident; thus a random-effects model was utilized.

Binary comparisons for the pooled mean difference analysis could only include participant characteristics of status (student v non-student) for Q1 and Q4. Assessment of all other variables (as listed in the correlation moderator analysis) could not be completed due to insufficient data (i.e., < 5 samples). Across both moderator analyses, effect sizes and variances were compared using the mixed effects analysis in CMA, with the Q statistic and $p < .05$ determining whether statistically significant differences existed between moderator binary comparisons.

2.2 Results

2.2.1 Systematic search

Figure 1 shows that 179,572 articles were identified in the systematic database search, along with searches in additional resources (e.g., reference lists of included papers). Applying the selection criteria subsequently led to 30 research studies being deemed eligible for systematic review inclusion. Of the 30, 21 (70%) reported a correlation between estimated and actual performance, while 23 (76.7%) described a trend of Q1 performers overestimating, and Q4's underestimating their performance relative to others (see column Q1 & Q4 in Table 1). Identified studies contained between 37 - 651,747 participants, with the overall total of participants involved in identified studies equalling 811,819. Mean ages across studies varied from 18-41, with 24.5-100% of samples being female. Participants were predominantly university students (70% of studies), completing tasks such as reasoning tasks (23.3%; e.g., logical reasoning), or various academic tests (16.7%; e.g., psychology exam).

Figure 2 summarizes the participant characteristics (e.g., age), methodological approach (i.e., timing of self-estimates, type of estimation scale, estimated performance reference, and actual performance comparison), proposed and/or measured influential psychosocial characteristics (e.g., narcissism), actual performance tasks (e.g., medical practice), manipulations and interventions (e.g., feedback), proposed DKE mechanisms (e.g., psychological based - metacognitive ability), as well as the potential individual consequences from low capabilities to align estimation with task performance. In total, the DKE has been examined in 33 different tasks, including for example: logical reasoning, math, humour, and gun safety. Within these contexts, various participant demographics, psychological characteristics, and study manipulations were shown (or hypothesized) to influence the capability to accurately match estimates with actual performance. Interestingly, while 6

Table 1: Illustrates studies identified from the systematic review process, presented chronologically and according to participant characteristics, task domain, performance test, DKE identification, and explanation(s) of findings.

Authors	N	F	Age	Participants	Domain	Performance Test	Q1	Q4	MA	Explanation
¹ Arsenian (1942)	125	un	un	Colleges students	Intelligence/english	ACE, cooperative english test	un	un	N	un
² Brim (1954)	103	un	un	Uni. students	Intelligence	ACE exam 1948 edi.	+	+	Y	un
³ Bailey & Lazar (1976)	40	50	un	Uni. students	Intelligence	Wechsler adults intelligence scale	un	un	Y	un
⁴ Kruger & Dunning (1999) _a	65	un	un	Uni. students	Humor	Self-made	+	+	Y	Metacognitive ability, skill level, false-consensus effect
⁴ Kruger & Dunning (1999) _b	45	un	un	Uni. students	Logical reasoning	LSAT	+	+	Y	Metacognitive ability, skill level, false-consensus effect
⁴ Kruger & Dunning (1999) _{cl}	84	un	un	Uni. Students	Grammar	NTE prep guide	+	+	Y	Metacognitive ability, skill level, false-consensus effect
⁴ Kruger & Dunning (1999) _d	140	un	un	Uni. students	Logical reasoning	Wason task	+	+	Y	Metacognitive ability, skill level, false-consensus effect
⁵ Furnham & Fong (2000) _{al-IV}	37	60.5	22.1	Uni. students	Intelligence	Raven's Standard Progressive Matrices (RPM)	un	un	Y	un
⁶ Haun et al. (2000)	41	un	un	Specimen-processing personnel	Specimen-processing	Self-made	+	+	N	Metacognitive ability, skill level, false-consensus effect
⁷ Krueger & Mueller (2002) _{al-II}	60	un	un	Uni. students	Grammar	NTE prep guide	+	+	Y	Regression-BTA
⁸ Ehrlinger & Dunning (2003) _a	59	74.6	un	Uni. students	Logical reasoning	LSAT	+	+	Y	Chronic self-views
⁸ Ehrlinger & Dunning (2003) _{bl-II}	91	70.3	un	Uni. students	Abstract reasoning/Computer programming	GRE	un	un	N	un

Table continues

Table 1 continued

Authors	N	F	Age	Participants	Domain	Performance Test	Q1	Q4	MA	Explanation
⁸ Ehrlinger & Dunning (2003) _{ci-II}	55	72.7	un	Uni. students	Geography	Self-made	un	un	N	un
⁹ Jaramillo et al. (2003)	172	un	un	Salespersons	Sales	Supervisor rating	+	+	Y	Self-presentation, modesty
¹⁰ Ames & Kammrath (2004) _a	143	24.5	28.4	Uni. students	Interpersonal sensitivity	IPT-15	+	+	Y	Skill level
¹⁰ Ames & Kammrath (2004) _{bi-II}	164	26.2	28.3	Uni. students	Interpersonal intentions/emotions	Self-made	+	+	Y	Skill level
¹¹ Furnham (2005) _{ai-aii}	100	69	18.2	Uni. students	Intelligence	Wonderlic personal test	un	un	Y	un
¹¹ Furnham (2005) _{bi-bii}	100	69	18.2	Uni. students	Intelligence	Baddeley reasoning test	un	un	Y	un
¹² Mikkelsson et al. (2005) _{ai}	64	55	40	40 Yr. olds	Endurance	V0 ₂ Max	un	un	Y	un
¹² Mikkelsson et al. (2005) _{aii}	64	55	40	40 Yr. olds	Speed	Counter-movement jump	un	un	Y	un
¹² Mikkelsson et al. (2005) _{aiii}	64	55	40	40 Yr. olds	Speed	Jumping in 15s	un	un	Y	un
¹² Mikkelsson et al. (2005) _{aiV}	64	55	40	40 Yr. olds	Strength	Sit-up test	un	un	Y	un
¹² Mikkelsson et al. (2005) _{aV}	64	55	40	40 Yr. olds	Strength	Hand-grip test	un	un	Y	un
¹² Mikkelsson et al. (2005) _{aVI}	64	55	40	40 Yr. olds	Flexibility	Sit-and-reach test	un	un	Y	un
¹³ Burson et al. (2006) _{ai-II, bi-VI}	130	un	un	Uni. students	Trivia	Self-made	+	+	Y	Noise-plus-bias
¹³ Burson et al. (2006) _{ci-IV}	76	un	un	Uni. students	English	Self-made	+	+	Y	Noise-plus-bias
¹⁴ Furnham et. al, (2006) _{a-c}	64	71.9	un	Uni. students	Intelligence	RPM, WPT, BRT	un	un	Y	un
¹⁵ Albanese et al. (2006)	113	un	un	Uni. students	Infection & immunity test	Academic test	+	+	N	un
¹⁶ Violato & Lockyer (2006) _{ai}	103	un	un	Specialist physicians	Patient management	PAR	+	+	N	Metacognitive ability, false-consensus effect
¹⁶ Violato & Lockyer (2006) _{aii}	103	un	un	Specialist physicians	Professional development	PAR	+	+	N	Metacognitive ability, false-consensus effect
¹⁶ Violato & Lockyer (2006) _{aiii}	103	un	un	Specialist physicians	Clinical assessment	PAR	+	+	N	Metacognitive ability, false-consensus effect
¹⁶ Violato & Lockyer (2006) _{aiV}	103	un	un	Specialist physicians	Ability to communicate	PAR	+	+	N	Metacognitive ability, false-consensus effect

Table continues

Table 1 continued

Authors	N	F	Age	Participants	Domain	Performance Test	Q1	Q4	MA	Explanation
¹⁶ Violato & Lockyer (2006) _{bl}	100	un	un	Paediatricians	Patient management	PAR	+	+	N	Metacognitive ability, false-consensus effect
¹⁶ Violato & Lockyer (2006) _{bII}	100	un	un	Paediatricians	Professional development	PAR	+	+	N	Metacognitive ability, false-consensus effect
¹⁶ Violato & Lockyer (2006) _{bIII}	100	un	un	Paediatricians	Clinical assessment	PAR	+	+	N	Metacognitive ability, false-consensus effect
¹⁶ Violato & Lockyer (2006) _{bIV}	100	un	un	Paediatricians	Ability to communicate	PAR	+	+	N	Metacognitive ability, false-consensus effect
¹⁶ Violato & Lockyer (2006) _{cl}	101	un	un	Psychiatrists	Patient management	PAR	+	+	N	Metacognitive ability, false-consensus effect
¹⁶ Violato & Lockyer (2006) _{cII}	101	un	un	Psychiatrists	Professional development	PAR	+	+	N	Metacognitive ability, false-consensus effect
¹⁶ Violato & Lockyer (2006) _{cIII}	101	un	un	Psychiatrists	Clinical assessment	PAR	+	+	N	Metacognitive ability, false-consensus effect
¹⁶ Violato & Lockyer (2006) _{cVI}	101	un	un	Psychiatrists	Ability to communicate	PAR	+	+	N	Metacognitive ability, false-consensus effect
¹⁷ Gross & Latham (2007)	51	73	18	Uni. students	Information literacy	ILT	+	+	Y	Metacognitive ability, skill level, false-consensus effect
¹⁸ Moore & Cain (2007) _{a, bl-II}	219	un	un	Uni. students	Easy/difficult trivia	Self-made	un	un	N	Differential regression
¹⁹ Moore & Small (2007) _{al}	124	50.8	un	Uni. students	Easy trivia	Self-made	+	+	Y	Differential regression
¹⁹ Moore & Small (2007) _{all}	131	61.1	un	Uni. students	Difficult trivia	Self-made	+	+	N	Differential regression
¹⁹ Moore & Small (2007) _{bl-II}	151	un	un	Uni. students	Predicting body weight	Self-made	+	+	N	Differential regression
²⁰ Ehrlinger et al. (2008) _a	124	un	un	Uni. students	Psychology	Academic test	+	+	Y	Skill level, false-consensus effect
²⁰ Ehrlinger et al. (2008) _b	58	un	un	Debaters	Debate	Judged Debate	+	-	N	Skill level, false-consensus effect
²⁰ Ehrlinger et al. (2008) _c	46	un	un	Trap & skeet competitors	Gun safety	NRA gun safety and knowledge test	+	+	Y	Skill level, false-consensus effect
²⁰ Ehrlinger et al. (2008) _d	57	un	un	Uni. students	Logical reasoning	Un	+	+	Y	Skill level, false-consensus effect

Table continues

Table 1 continued

Authors	N	F	Age	Participants	Domain	Performance Test	Q1	Q4	MA	Explanation
²⁰ Ehrlinger et al. (2008) _e	42	un	un	Uni. students	Logical reasoning	LSAT	+	+	Y	Skill level, false-consensus effect,
²¹ Battistelli et al. (2009)	65	100	un	Uni. students	Linguistic, mathematical, & logical reasoning	Self-made	+	+	N	optimism, methodical doubt
²² Ferraro (2010) _{aI-aIII}	105	58.1	un	Uni. students	Economics test	Academic test	+	+	Y	un
²² Ferraro (2010) _{bI-bIV}	64	un	un	Uni. students	Economics test	Academic test	+	+	Y	un
²³ Mattern et al. (2010) _a	153, 961	53%	un	Uni. students	Math	SAT	+	+	Y	Skill level
²³ Mattern et al. (2010) _b	651, 747	56%	un	Uni. students	Math	SAT	+	+	Y	Skill level
²⁴ De Craen et al. (2011) _{aI-II}	83	48	20	Novice Dutch drivers	Driving	On-road driving assessment	+	+	Y	un
²⁴ De Craen et al. (2011) _{aIII-IV}	47	51	41	Experienced Dutch drivers	Driving	On-road driving assessment	+	+	Y	un
²⁵ Gross & Latham (2012) _{aI-II}	287	58.9	21.2	College students	Information literacy	ILT	+	-	Y	Recalibration inability, metacognitive ability, skill, false-consensus effect
²⁵ Gross & Latham (2012) _{bI-II}	290	52.4	20.6	College students	Information literacy	ILT	+	+	Y	Recalibration inability, metacognitive ability, skill, false-consensus effect
²⁶ Ryvkin et al. (2012) _{aI-III}	58	30	26	Uni. students	Microeconomics test	Academic test	+	+	Y	Self-presentation, informational asymmetry
²⁶ Ryvkin et al. (2012) _{aIV-VI}	53	49	24	Uni. students	Microeconomics test	Academic test	+	+	Y	Self-presentation, informational asymmetry

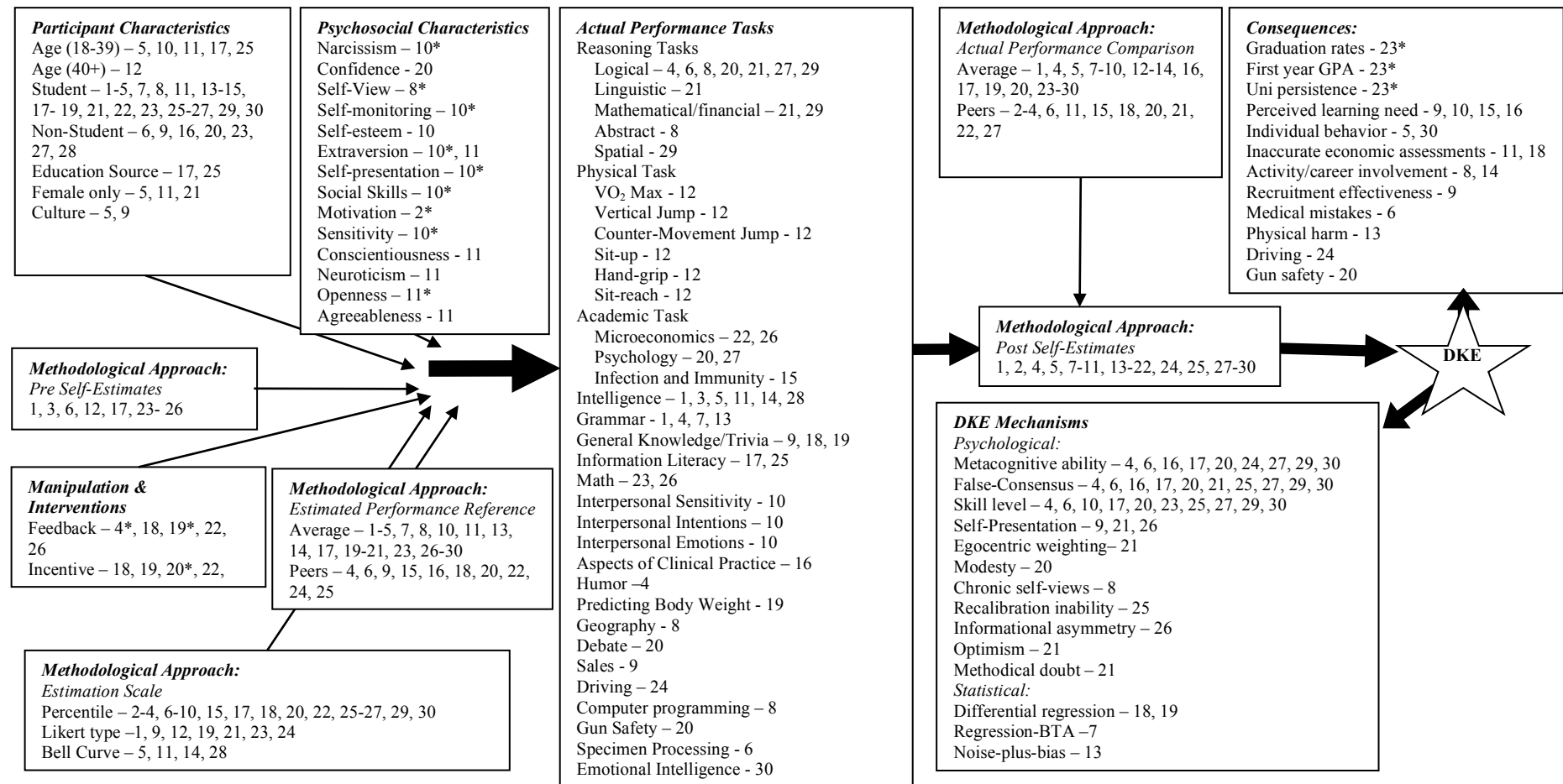
Table continues

Table 1 continued

Authors	N	F	Age	Participants	Domain	Performance Test	Q1	Q4	MA	Explanation
²⁶ Ryvkin et al. (2012) _{bl-IV}	58	30	26	Uni. students	Math/Geography	Self-made	+	+	Y	Self-presentation, informational asymmetry
²⁶ Ryvkin et al. (2012) _{bV-VIII}	53	49	24	Uni. students	Math/Geography	Self-made	+	+	Y	Self-presentation, informational asymmetry
²⁷ Schlösser et al. (2013) _b	344	un	un	Uni. students	Psychology	Academic test	+	+	Y	Metacognitive ability, skill level, false-consensus effect
²⁷ Schlösser et al. (2013) _c	103	un	un	General population	Logic	Wason task	+	+	Y	Metacognitive ability, skill level, false-consensus effect
²⁸ von Stumm (2013)	176	52	35	General population	Intelligence	RPM, various IQ tests	+	+	Y	un
²⁹ Williams et al. (2013) _a	140	un	un	Uni. Students	Logical reasoning	Wason task	+	+	Y	Metacognitive ability, skill level, false-consensus effect
²⁹ Williams et al. (2013) _b	102	un	un	Users of Mechanical Turk	Spatial reasoning	Self-made	+	+	Y	Metacognitive ability, skill level, false-consensus effect
²⁹ Williams et al. (2013) _c	102	un	un	Users of Mechanical Turk	Financial reasoning	Self-made	+	+	Y	Metacognitive ability, skill level, false-consensus effect
²⁹ Williams et al. (2013) _{d-f}	339	un	un	Users of Mechanical Turk	Logical Reasoning	Wason task	+	+	Y	Metacognitive ability, skill level, false-consensus effect
³⁰ Sheldon et al. (2014) _{a-cll}	364	47	29	Uni. students	Emotional Intelligence	MSCEIT	+	+	Y	Metacognitive ability, skill level, false-consensus effect

Table Key: Superset numbers before author citation refer to chronological order of the study; subset letter refers to macroscopic classification of independent study within the article (e.g., a = study 1, b=2 etc); roman numerals identify microscopic classification of specific sub-set information (e.g., different task) in an independent study within an article; N = number of participants in study; F = percentage of females in the sample; un = unknown values; Age = average age of participants, Domain = skill/capability area examined; Performance Test = specific test used to assess skill/capability; Q1 = assessment of whether overestimation (i.e., DKE effect) was found in the lowest performing quartile; Q4 = assessment of whether underestimation (i.e., DKE effect) was found in the highest performing quartile; MA = whether data was available for meta-analysis; explanation = account provided by authors to explain their findings; + = significant over/under estimation found, - = non-significant over/under estimation found; Y = data available for meta-analysis; N = no data available for meta-analysis.

Figure 2: Summary chart overviewing participant demographic, psycho-social characteristics, methodological approaches, actual performance tasks, individual consequences, & mechanisms associated with the DKE.



Note: Numbers refer to corresponding numbered articles in the reference list. - = range of articles. * = correlated with/alterd estimated performance.

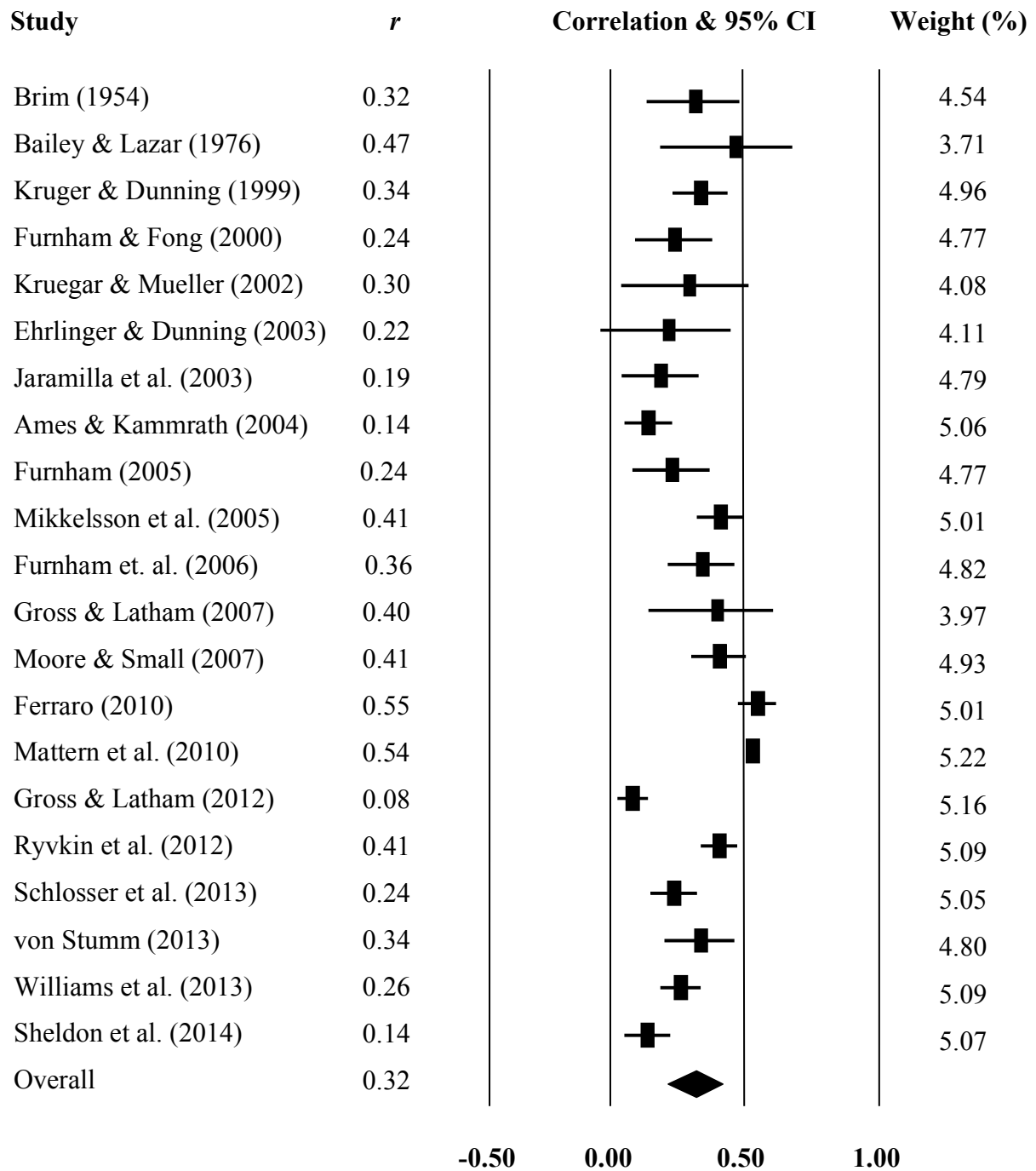
studies were designed to intervene or mitigate the DKE, 14 proposed DKE consequences, and only 1 provided evidence of specific estimation-performance consequences (see Figure 2). In terms of methodological approach, 9 (30%) studies asked participants to self-estimate prior to the task, 25 (83.3%) asked post task, and 5 (16.7%) asked for estimates both prior and post. To obtain estimates, 18 (60%) used a percentile ranking scale, 7 (23.3%) used a Likert ranking scale, while 4 (13.3%) adopted a bell curve distribution. Across all the performance tasks, 6 (18.2%) were classified as physical, 22 (66.7%) as cognitive (with 3 or 10% classified as academic within this category), and 5 (15.2%) as other. A consistent methodology (i.e., same comparison group for estimation & actual performance) was used by 66 (89.2%) samples within the studies, whilst the remaining 8 (10.8%) were inconsistent (i.e., different comparison group for estimation & actual performance). Sixty (81.1%) samples had participants estimate relative to an ‘average’ population and compared actual performance to the ‘average’; while 6 (8.1%) estimated relative to peers but then compared performance to the ‘average’; 6 (8.1%) estimated relative to peers and compared to their peers, and 2 (2.7%) of samples estimated relative to the ‘average’, but compared performance to their peers.

2.2.2 Meta-analysis

2.2.2.1 Overall

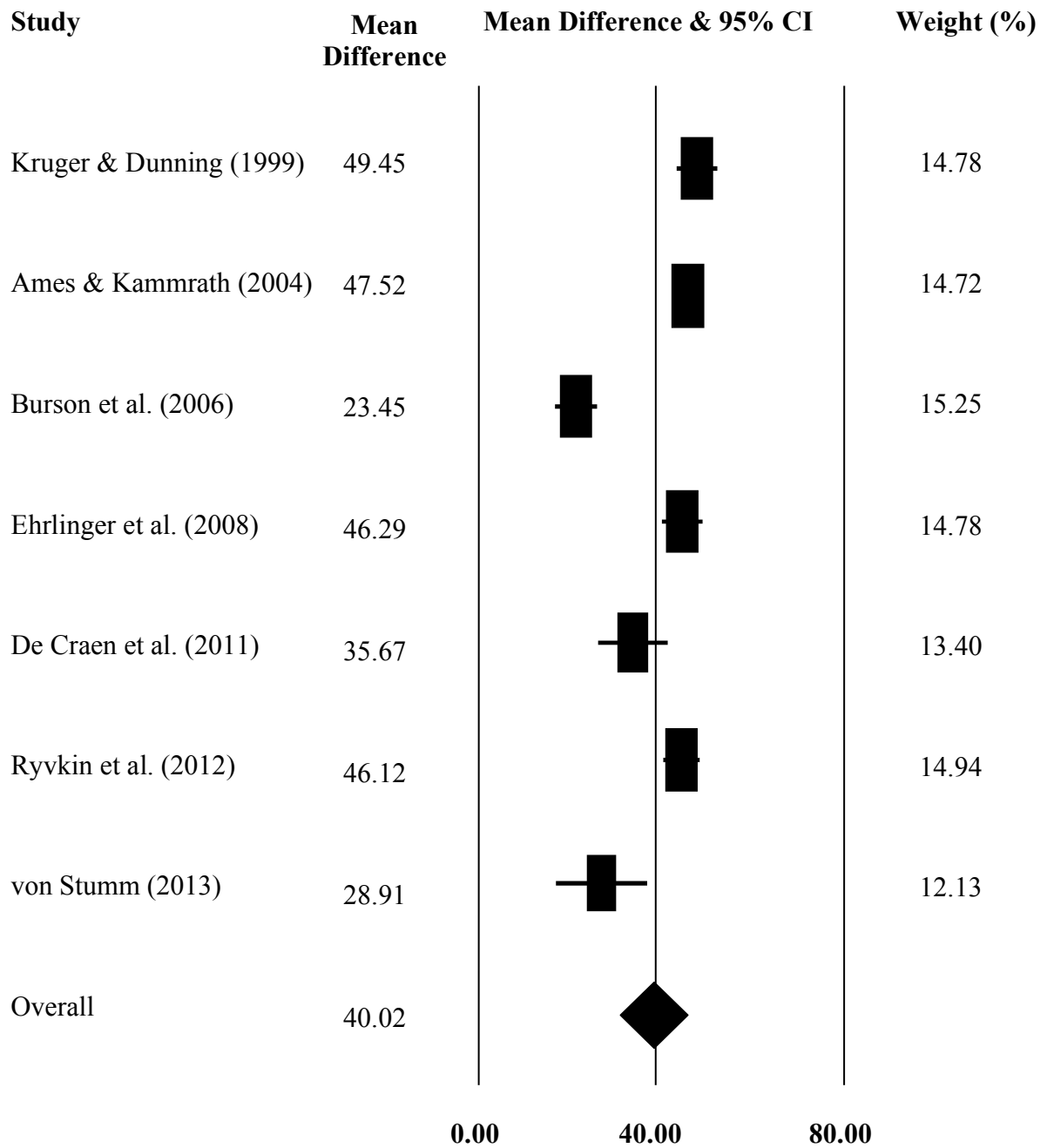
A total of 23 studies were taken forward for meta-analysis, among which 21 studies (74 samples) were included in the pooled correlation analysis, and 7 studies (39 samples) included in the DKE pooled mean difference analyses. The overall pooled correlation between estimated and actual task performance identified a small-medium correlation ($r = .32$, $CI = 0.29 - 0.35$, $z \text{ value} = 20.94$, $p < .001$; see Figure 3), indicating a small-moderate ability of individuals to accurately estimate performance. In comparing mean percentile differences (i.e., estimated vs. actual performance) in Q1 performers, there was a strong pooled effect size with a stable

Figure 3: Forest plot of correlations between estimated and actual relative performance across and within studies.



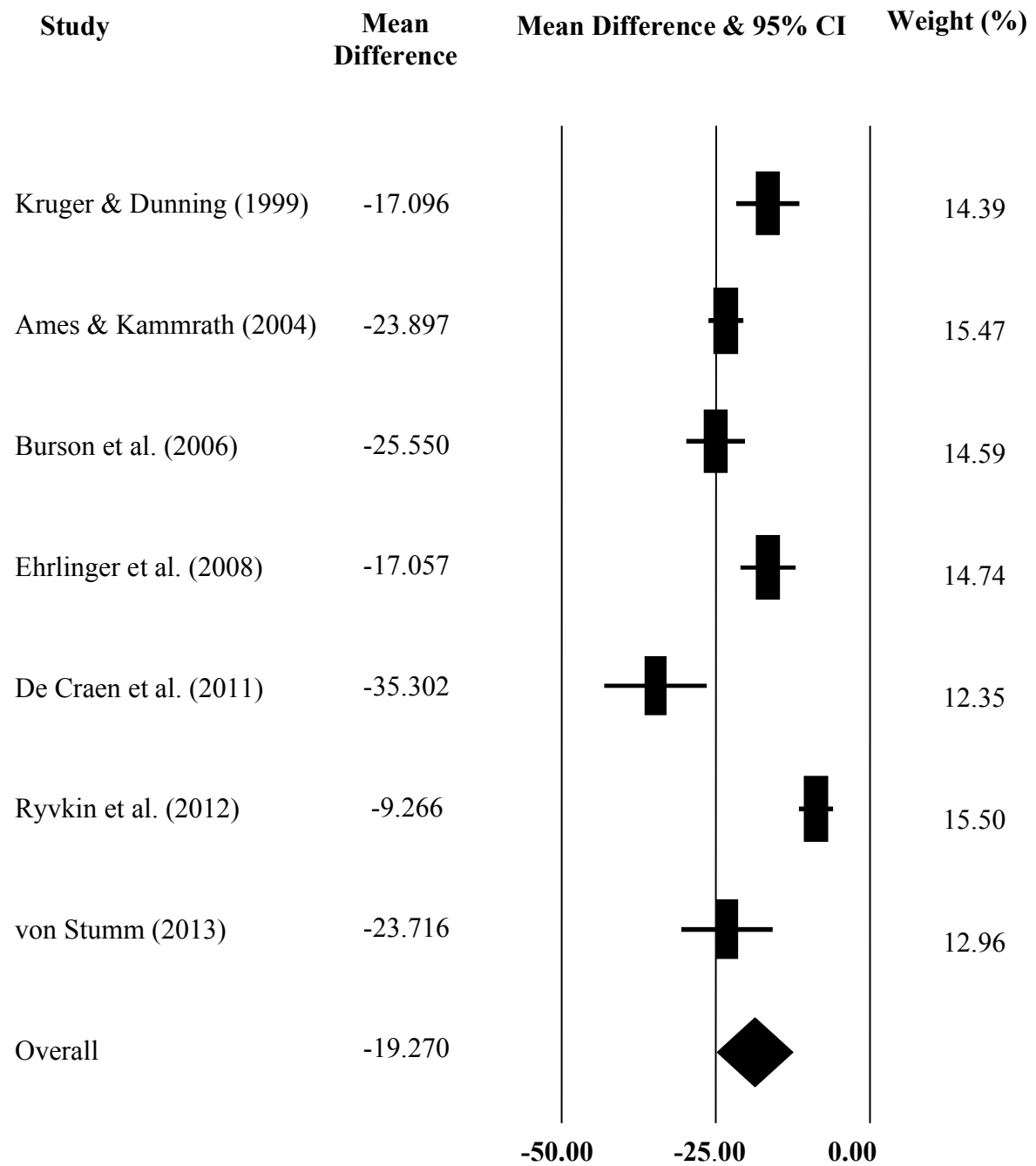
Note: Several studies above contain additional samples/independent sub-studies within them. In this figure, for illustration purposes, samples are combined according to study. *r* = Pearson's product correlation coefficient; CI = confidence interval; Weight (%) = percent of contribution to overall correlation by each study.

Figure 4: Forest plot of quartile one (Q1) performer’s overestimation across and within studies.



Note: Several studies above contain additional samples/independent sub-studies within them. In this figure, for illustration purposes, samples are combined according to the study. CI = confidence interval; Weight (%) = percent of contribution to overall correlation by each study.

Figure 5: Forest plot of quartile four performer's (Q4) underestimation across and within studies.



Note: Several studies above contain additional samples/independent sub-studies within them. In this figure, for illustration purposes, samples are combined according to the study. CI = confidence interval; Weight (%) = percent of contribution to overall correlation by each study.

and consistent trend of overestimation (PE = 37.44, CI = 33.37 - 41.51, $z = 18.04$, effect size $[g] = 2.17$, $p < .001$; see Figure 4); while for Q4 performers, there was also a consistent and strong pooled effect size for underestimation (PE = -19.96, CI = -23.34 - -16.58, $z = -11.58$, $g = -1.22$, $p < .001$; Figure 5). In comparing Q1 and Q4 performers estimations, a significant difference in the magnitude of mis-estimation was found (i.e., $Q = 452.67$, $p < .001$), with Q1's consistently higher in estimation error.

2.2.2.2 Correlation moderator analyses

When accounting for variations of methodological approaches and task contexts examined between studies, significant moderation of correlations were evident. These specifically related to the timing of performance estimates (pre, $r = .43$, CI = 0.40 - 0.47, $z = 21.13$, $p < .001$; post, $r = .30$, CI = 0.26 - 0.35, $z = 11.51$, $p < .001$; pre v post $Q = 17.94$, $p < .001$), the type of performance measurement scale used (Likert, $r = .50$, CI = 0.47 - 0.52, $z = 27.75$, $p < .001$; percentile, $r = .30$, CI = 0.25 - 0.36, $z = 9.82$, $p < .001$; bell curve, $r = .30$, CI = 0.23 - 0.37, $z = 8.07$, $p < .001$; Likert v percentile $Q = 37.59$, $p < .001$; Likert v bell curve $Q = 28.93$, $p < .001$) and for methodological consistency (i.e., consistent comparisons across estimated & actual performance, $r = .36$, CI = 0.34 - 0.39, $z = 24.86$, $p < .001$; inconsistent comparisons across estimated & actual performance, $r = .16$, CI = 0.07 - 0.25, $z = 3.33$, $p = .001$; consistent v inconsistent methodology $Q = 19.04$, $p < .001$). Finally, differences in performance task type (physical, $r = .41$, CI = 0.32 - 0.50, $z = 8.21$, $p < .001$; cognitive, $r = .32$, CI = 0.29 - 0.34, $z = 20.94$, $p < .001$; physical v cognitive $Q = 4.20$, $p = .040$) were also apparent. Together, these findings indicate that estimation-performance correlations were lowered (i.e., became less aligned) when participants were asked to estimate performance after task completion, when using a percentile or bell-curve scale, with dissimilar estimation reference and performance comparison groups, and when performing cognitive task types.

When studies applied a consistent methodology for estimation and performance reference comparisons (e.g., estimation relative to ‘average’ & actual performance relative to ‘average’, $r = .35$, $CI = 0.33 - 0.38$, $z = 22.78$, $p < .001$; estimation relative to ‘peers’ & actual performance relative to ‘peers’, $r = .48$, $CI = 0.34 - 0.61$, $z = 5.83$, $p < .001$), compared to when inconsistent (e.g., estimation relative to ‘peers’ & actual performance relative to ‘average’, $r = .11$, $CI = 0.03 - 0.19$, $z = 2.63$, $p = .010$), higher correlations were observed. However, insufficient data meant that determination of significance could not be made.

Correlations were not affected by the participant characteristics of gender (female $r = .32$, $CI = 0.23 - 0.41$, $z = 6.53$, $p < .001$; males $r = .35$, $CI = 0.25 - 0.45$, $z = 6.20$, $p < .001$; female v male $Q = 0.25$, $p = .620$), age (i.e., participants ≥ 40 years $r = .41$, $CI = 0.32 - 0.50$, $z = 8.21$, $p < .001$; < 40 years $r = .36$, $CI = 0.26 - .45$, $z = 6.50$, $p < .001$; < 40 years v ≥ 40 years $Q = 0.70$, $p = .404$), and participant status (students $r = .33$, $CI = 0.30 - 0.36$, $z = 20.53$, $p < .001$; non-students $r = .29$, $CI = 0.24-0.34$, $z = 11.07$, $p < .001$; student v non-student $Q = 1.81$, $p = .180$). Likewise, they were not affected by particular study methodological factors, such as when estimating performance using percentiles or a bell-curve (i.e., percentile v bell curve, $Q = 0.00$, $p = .957$), or particular task type (i.e., academic, $r = 0.52$, $CI = 0.40-0.62$, $z = 7.44$, $p < .001$; physical v academic, $Q = 2.09$, $p = .148$). These findings indicate effect robustness, and that basic socio-demographic participant characteristics do not affect estimation-performance correlation (in)capability, along with particular estimation scales and task contexts.

2.2.2.3 Mean difference moderator analyses

Significant moderation in mean differences between estimated and actual performance were apparent for Q4 performers (i.e., students, $PE = -18.13$, $CI = -21.62 - -14.64$, $z = -14.64$, $g = -1.19$, $p < .001$; non-students, $PE = -31.15$, $CI = -40.42 - -21.88$, $z = -6.59$, $g = -1.34$, $p < .001$; students v non-students, $Q = 6.64$, $p = .010$). Specifically, Q4

students underestimated (i.e., -18.13 percentile points) their relative performance less than Q4 non-students (i.e., -31.15 percentile points), suggesting that Q4 non-students (typically older individuals) consistently made substantially lower performance estimates, possibly reflecting more caution or conservative estimation. Participant status did not moderate estimation-performance correlations in Q1 performers, with non-students overestimating to a similar magnitude of students (i.e., non-students, $PE = 34.55$, $CI = 25.64 - 43.46$, $z = 7.60$, $g = 1.45$, $p < .001$; students, $PE = 37.96$, $CI = 33.56 - 42.36$, $z = 16.91$, $g = 2.32$, $p < .001$; $Q = 0.45$, $p = .501$).

2.3 Discussion

2.3.1 Meta-analysis findings

This meta-analysis aimed to determine a quantifiable pooled (across study) correlation estimate for the capability to align estimation with performance; and uniquely in terms of DKEs, determine an overall pooled mean difference between estimated-actual performance estimates for both Q1 and Q4 performers. Secondly, the analysis of moderating factors was conducted to identify whom (i.e., types of participants), what situations/contexts (i.e., tasks), and what methodological factors were consistently related to both lower/higher correlations and DKE over/under estimations. Findings identified a significant pooled correlation estimate with a small-moderate effect size based on 21 predominantly equally weighted studies and 74 within-study samples, similar to partially related meta-analyses (e.g., Freund & Kasten, 2012). This reflects a consistent limited overall capability to accurately align self-estimates with actual performance when comparing oneself relative to others, whether it is class peers or a broader population, and regardless of participant, task, and methodological factors and inconsistencies.

In terms of pooled mean difference estimates, meta-analytical findings are original, identifying consistent and significant differences for Q1 and Q4 performers based on 7 evenly weighted studies that contained 39 independent samples. This confirms DKE's prevalence, consistency, and relevance as an explanation for low estimation-performance calibration, as low (Q1) and high (Q4) competency performers show inaccuracy in opposing directions. That is, Q1's overestimate, and generally by a higher magnitude of estimation error, relative to Q4 underestimation. Therefore, the combination of the pooled correlation and pooled mean difference analyses indicate that individuals are consistently unable to accurately estimate performance. With the top and bottom performers being the most inaccurate.

Further original and important findings arose from the moderator analyses, which highlighted that sources of heterogeneity in study results, to a greater extent, were related to key systematic methodological factors. For instance, the timing of self-estimates (i.e., pre v post-task estimate) significantly affected estimation-performance correlations with post-task estimates lowering overall alignment. Although this contradicts convention that experience should increase estimation-performance accuracy, pre-task estimation may be more indicative of actual performance due to the assimilation of prior performance (Ehrlinger & Dunning, 2003). Post-task estimates however may be more influenced by psycho-social characteristics such as confidence (Felson, 1981), decreasing accuracy.

Supporting the claim made by Fitzgerald et al. (2003), task context also significantly affected estimation-performance correlations, with alignment in cognitive tasks being lower than in physical tasks. The finding can again be associated with DKE mechanisms, as cognitive tasks may have been less known and familiar to study participants compared to physical tasks examined. That is, task familiarity in physical contexts may have permitted improved estimation-performance accuracy (Fitzgerald et al., 2003; Mullen et al., 1985).

However, Kruger and Dunning (1999) suggested that in cognitive task situations where similar specific forms of knowledge are necessary to both perform and evaluate the skill, a lack of metacognitive skill will lead to decreases in estimation-performance accuracy, whereas in physical tasks, the skills necessary for performance (e.g., running), can be dissimilar to those needed for evaluation (e.g., subjectively or objectively measuring technique or speed). It is also likely that estimation-performance accuracy may be better in physical contexts, as individual differences may be more observable, or easier to identify explicitly, regardless of actual performance experience. Thus metacognitive skill is pertinent not only to cognitive tasks, but also seems to be critical for estimation-performance alignment regardless of context (Keith & Frese, 2005; Schmidt & Ford, 2003; Schraw, 1998).

The type of estimation scale applied in study methodological procedures significantly moderated estimation-performance correlations. Higher correlations occurred when participants estimated performance on a Likert scale, compared to a percentile or bell-curve scale. Likert scales typically have a limited number of possible responses (i.e., 5-7), compared to the responses available using a percentile or bell-curve scale (e.g., 100), and so Likert scales can inflate estimation-performance accuracy by increasing the probability of estimating correctly (i.e., a ‘false positive’). If study participants were truly better in estimation-performance alignment, then increasing estimation specificity via percentile scales should not have affected the moderation analysis.

Another expected finding was that estimation-performance correlations were moderated by methodological inconsistencies regarding how estimates and reference comparisons were made (e.g., estimations relative to peers or the average of a broader population). Logically, if there were a discrepancy between the reference groups used for estimation and actual performance assessment, then decreases in estimation-performance alignment would be apparent. However, when studies asked for performance estimates

relative to peers, estimation-performance correlations improved compared to when estimations were requested relative to a broader or average population group. This indicates that reference group familiarity or proximity influences the capability to align estimations with performance.

Consistent with individual studies (Ames & Kammrath, 2004; Ehrlinger & Dunning, 2003; Mikkelsen et al., 2005), pooled estimation-performance correlations were not moderated by the participant characteristics of gender, status, and age. These findings (or lack thereof) are informative, as they support the notion that estimation-performance alignment reflects a learned and acquired quality, associated with skill competency and metacognitive skill, as opposed to characteristics of individuals per se (Flavell, 1979). This indicates individuals can learn to become more accurate in their self-assessments, versus being unable to influence their accuracy. This is also supported by pooled mean difference moderations according to participant status. Specifically, participant status had no relationship with Q1 performers; but in Q4 students, young student self-assessments were more accurate than the typically older non-students. Many of these studies used academic tasks for testing, and it is feasible that Q4 students benefitted from increased task familiarity. Lower task familiarity and greater metacognitive knowledge and awareness may have more invoked the false-consensus effect (Fitzgerald et al., 2003; Mullen et al., 1985) in non-students, reflected by overly conservative estimations (i.e., greater underestimation).

As supported by individual studies (e.g., Gross & Latham, 2012; Schlösser et al., 2013; Sheldon et al., 2014; Williams et al., 2013) and when considered together, meta-analytical findings can be accounted for by Kruger and Dunning's (1999) DKE explanations. Chiefly, they explain why some individuals are less capable than others of aligning estimation with performance accurately, and while alternative mechanisms have been proposed (e.g., statistical constructs; Burson et al., 2006; Krueger & Mueller, 2002), a

growing body of literature has addressed these concerns providing further validity to DKE explanations (Albanese et al., 2006; Ehrlinger et al., 2008; Kruger & Dunning, 2002).

2.3.2. Findings from systematic reviewing

Through systematic reviewing of identified studies and data extraction, additional complementary findings were identified. These are summarised in table and flow-chart formats (see Table 1 & Figure 2) where the breadth and focus of related studies can be seen. For instance, Figure 2 shows studies which have examined (or hypothesized) the influence of psycho-social characteristics, the consequences of estimation-performance incapability and DKEs, and interventions that have attempted to remove or improve such estimation-performance (in)capability. These were considered and integrated alongside meta-analysis findings.

2.3.2.1 Psycho-social characteristics

Several studies have reported (or theorized) that psycho-social characteristics influence estimation-performance alignment. Influencing overall performance estimations, these can include narcissism, self-monitoring, sensitivity to others, self-presentation, extraversion, neuroticism, openness to experience, conscientiousness, agreeableness, self-view, and social skills (Ehrlinger & Dunning, 2003; Gati, Fishman-Nadav, & Shiloh, 2006); with narcissism (Ames & Kammrath, 2004), and openness to experiences (Furnham, 2005) increasing estimations of performance regardless of performance quartile. Influencing specific quartiles for example, is egocentrism in Q1 performers (e.g., Merkle & Weber, 2011) and modesty in Q4 performers (e.g., Ehrlinger et al., 2008), further exaggerating estimation error. It is proposed that a benefit of metacognitive skill however, may be to actively mitigate these potential psycho-social characteristics, decreasing their influence on overall estimations and quartile specific estimation error. Individuals with low competency and metacognitive skill (i.e., Q1's) however, may not be able to acknowledge this influence and compensate,

thus rendering them more vulnerable to psycho-social bias and estimation-performance mal-alignment.

2.3.2.2 Consequences

DKE mechanisms can be used to explain how estimation-performance inaccuracies lead to detrimental consequences, as highlighted by limited studies (e.g., dismissed from university; Arsenian, 1942). On a positive note initially, Q1 increases in perceived capability (although false) may heighten motivation to continue specific task involvement (Cury, Biddle, Sarrazin, & Famose, 1997). However, a more concerning matter is that Q1 overestimation, due to metacognitive deficits, has been predicted to lead to numerous consequences. These may include: increased health risk (Lee, 1989); financial loss (Ferraro, 2010); property and personal loss (De Craen et al., 2011; Jonah, 1986); patient death (Whitaker, 2008); individual injury and death during recreational activities (Burson et al., 2006; Palmer, 2002; Petrass et al., 2012); decreased recruitment effectiveness and perceived need for training in sales personnel (Jaramillo et al., 2003); medical laboratory mistakes (Haun et al., 2000); inappropriate weapon safety (Ehrlinger et al., 2008); and deviant driving behaviour (De Craen et al., 2011).

Studies have also identified consequences via Q4 performers. For example, when compared to more accurate 'top performing' students starting university, Mattern et al. (2010) found that underestimating 'top performing' students attained a lower GPA (i.e., by 0.4/4.0) by the end of their first year; were less likely to persist into the fourth year of university (i.e., a 11% decrease) and then graduate from university (i.e., a 18% decrease). Further, Q4's lack of perceived performance capability may lead to a reduction in continued task involvement (Cury et al., 1997), forgoing opportunities and continued learning (e.g., women forgoing science careers, Ehrlinger & Dunning, 2003), and ultimately leading to a decrease in performance capability.

2.3.2.3 Interventions

In order to decrease estimation-performance error, some studies have increased the general feedback available to participants. This has been done through providing post task descriptive statistics (Ferraro, 2010), performance rank (Moore & Cain, 2007), and absolute or relative performance information (Ryvkin et al., 2012). Other studies have used various incentives in an effort to encourage more accurate estimations (e.g., monetary reward; Ehrlinger et al., 2008; Ferraro, 2010). Unfortunately, studies using general feedback and incentive approaches have been largely unsuccessful, with monetary incentives exacerbating estimation-performance error by 25-31 percentile points (Ehrlinger et al., 2008). Instead, only studies that have purposefully targeted metacognitive skill and false-consensus mechanisms have seen significant improvement in estimation accuracy.

In targeting metacognitive skill, Kruger and Dunning (1999; see study 4) had university students initially complete a logical reasoning task, and then estimate their performance on it. Half then completed a 10-minute ‘training packet’ intervention which highlighted methods for detecting the accuracy, or flaws in, logical syllogisms (i.e., knowledge & metacognitive skill); while the other half completed a ‘filler task’. All participants then reviewed their test and re-estimated performance. Intervention exposure decreased Q1 overestimation by 18.6 percentile points ($p < .001$) and Q4 overestimation by 13.4 percentile points ($p < .05$), while no changes occurred in controls.

In targeting the false-consensus effect, Kruger and Dunning (1999; see study 3) invited Q1 and Q4 performers to return and review their performance in a grammar test, and grade five prior peer-performed tests, before re-estimating their initial performance. A decrease in estimation error by Q4 performers was found (i.e., 5.6 percentile points; $p < .010$), while Q1 performers showed no improvement.

2.3.3 Recommendations and future directions

As methodological variation and inconsistencies were shown to influence correlation analyses, a primary recommendation is that researchers rigorously consider their methodological procedures, forms of measurement for estimation and performance, and consistency with respect to reference or comparison groups. This recommendation is not intended to homogenize research methodology, but rather to encourage the proper alignment of study purpose and methodology to ensure overall alignment of study goals, methods, results, and discussion. Additional recommendations arising from this review include asking for estimations pre-task, the use of percentile ranks (or bell-curves), consistency in applying scales during estimation and performance, and using consistent reference groups (e.g., similar age-matched peers). These would help ensure that researchers and participants can most accurately capture and provide estimations respectively without unwanted influence of extraneous factors.

For both researchers and practitioners, it would be valuable to establish whether particular psycho-social characteristics (see Figure 2) are influential to estimation-performance capabilities, and whether metacognition can mediate this relationship. Likewise, identifying the behavioural, cognitive, learning and health consequences of (consistent) estimation-performance error in specific tasks would provide much impetus to the research area. Finally, implementing interventions that aim to improve overall estimated-performance accuracy, and simultaneously reduce errors in Q1 and Q4 performers will be informative, particularly if clear consequences of estimation-performance are confirmed.

2.3.4 Limitations

Key limitations in this study relate mainly to data availability. Data extraction was often limited by the absence or non-reporting of data within studies. Frequently, required data (e.g., participant gender and age range) was not provided, and so could not be included in

pooled correlations, pooled mean differences, or moderator analyses. To add, due to prioritization of identifying meta-analytical mean difference trends in Q1 and Q4 performers, analyses of Q2 and Q3 performers were not conducted.

While the meta-analysis followed PRISMA guidelines throughout, risk of bias was not investigated within the individual studies. While a lack of overall publication bias was determined, studies were deemed suitable based solely on meeting inclusion criteria, and not on more subjective criteria of quality. Due to the scope and purpose, this meta-analysis restricted itself to the literature that investigated self-assessment accuracy in relation to the DKE and actual performance. Therefore research investigating more general self-assessments, although potentially informative regarding metacognition and self-assessment, were left out. While this allowed for a specific and informative discussion, this limited the generality of the results, and may have led to the omission of possible explanative factors regarding the DKE.

Additionally, due to the number and nature of the included studies, the meta-analysis did not attempt to contrast the various mechanisms of the DKE. Instead it relied on the results of its included studies to determine which of the proposed mechanisms explains the DKE. Lastly, while the meta-analysis found significant support for the DKE and its mechanisms in self-assessment contexts, limited studies directly investigated metacognition in participants. While results align with the DKE, the mechanisms behind it are more suspect. Overall, while the meta-analysis is able to provide substantial evidence for the DKE, these limitations restrict the full acceptance of its metacognitive mechanism.

Chapter 2 - 3 Bridging Statement

Chapter 2 determined the historical and present scope of research associated with the ability to align estimation and performance, specifically in terms of the DKE. The review enabled a detailed account of what has been investigated in the research area, and helped determine what is not known, what remains uncertain, and what research directions and questions still need to be examined. For instance, while the explanations for correlations, pooled mean differences, and moderators of estimation accuracy have provided significant insight into the cognitive processes responsible for such trends, the existence of DKEs in non-cognitive contexts (e.g., physical tasks) have not been determined. Further, methodological inconsistencies within existing studies, such as when the timing of self-estimates are made (i.e., pre v post), may explain variable findings and contexts where particular cognitive processes are (not) at work. The influence, control, and/or manipulation of variables such as participant task experience, task familiarity, performance feedback, and task difficulty on individual self-assessment accuracy have been insufficiently examined; yet may also account for the nature and magnitude of estimation-performance misalignment.

Chapter 3 thus aimed to address some of these unexplored questions in lesser-studied contexts, by examining self-assessment error in the DKE and its underlying processes, in two common physical activity contexts (i.e., Sprint and Vertical Jump). Likewise, the potential confounding influence of previous task experience was considered, while manipulation occurred for when the timing of performance estimations was made (i.e., pre v post task), and the reference group against who estimations were made (i.e., similar aged students v athletes). This study therefore attempted to provide an informative and reliable examination of DKE prevalence in physical activity contexts, determine whether task experience could potentially explain DKE trends, as well as examine the cognitive processes related to self-assessment and estimation-performance (mal)alignment. A better understanding of how

individuals utilize, for instance, prior experience, chronic self-views and performance feedback to inform perceptions of performance rank would assist understanding how detrimental consequences of self-assessment inaccuracy occur, and highlight potential avenues for intervention. This would then lead to significant contribution to the mitigation of estimation inaccuracies, allowing individuals to consistently and accurately self-assess.

Chapter 3 – The Capability to Align Estimated and Actual Performance: The Dunning-Kruger Effect is Evident in Physical Task Contexts

Abstract

Background: The Dunning-Kruger effect (DKE) demonstrates how both the competent and incompetent are unable to accurately align task estimation with actual performance. **Objectives:** This study aimed to determine DKE validity and reliability in the common physical tasks of Sprint and Vertical Jump, assessing whether sporting experience, estimation time point, and reference group moderated estimated performance. **Methods:** The relationship between participant's actual performance was compared to their estimated performance, while determining their relationship with the various potential moderators. Therefore correlations, *t*-tests, and an ANOVA were used to fully detail the relationship. **Results:** Across both tasks typical DKEs were apparent, with significant overestimation in Quartile 1 'low performers' (i.e., 31 to 35 percentile points) and underestimation in Quartile 4 'high performers' (i.e., -30 to -29 percentile points). Estimation error was not related to current or previous sporting experience in either task. Sprint estimation rank was affected by time point, with greater error occurring following task attempts (v pre-test), while estimates in both tasks were decreased when an athlete (v student population) reference comparison was used. **Conclusion:** DKE prevalence was confirmed, with potential implications for learning and participation in sport and physical activity.

Keywords: Cognitive Bias, Self-Assessment and Evaluation, Task Competence

3.0 Introduction

While DKEs seem prevalent, as discussed in Chapter 2, several methodological incongruities have led to variable findings, raised questions regarding DKE's existence in both broad and specific contexts; and generated debate about DKE's etiology. For instance, a high proportion of studies assessed academic or cognitive tasks, with less assessment of non-academic (e.g., physical) or more common day-day tasks. Thus, determining DKE's ecological reach is important. In existing studies, participants have been less knowledgeable or experienced in respective tasks compared to typical performance contexts, with ongoing or immediate feedback not permitted. Participants have also been asked to make inconsistent performance estimates relative to various reference groups (e.g., peer v average population; De Craen et al., 2011), either before (e.g., Ryvkin et al., 2012) or after task performance (e.g., Battistelli et al., 2009).

While studies have sought to improve estimation accuracy through the use of increased performance feedback, these interventions have mainly been conducted after performance (e.g., Ryvkin et al., 2012). Increasing the amount of performance feedback individuals can receive during performance has been suggested to improve metacognitive skill (Butler, Karpicke, & Roediger III, 2008), which has in turn has been shown to improve estimation accuracy (Kruger & Dunning, 1999). Additionally, an increased task difficulty (Burson et al., 2006), coupled with lower task familiarity and skill (Cox & Griggs, 1982; Kruger & Dunning, 1999), has been highlighted as having the potential to decrease overall performance estimates.

Investigating individual performance estimation during physical tasks, ones that are easy to perform, familiar, and which provide a large degree of ongoing performance is therefore important. This will not only allow for the determination of the DKEs validity in

unexplored situations, but also determine potential methods of reducing individual estimation error throughout individual performance.

Thus on these premises, the general purpose of this study was to determine the capability of individuals to accurately align performance estimations with actual performance. More specifically, it aimed to determine the correlations and mean differences between actual and estimated performance according to performance quartile on two separate, yet familiar and common physical activity tasks. Therefore, the first prediction of this study was that, participants would display higher estimation accuracy as compared to previous studies in the literature.

This study also sought to determine the extent of quartile specific estimation error (i.e., direction & magnitude) between their estimated and actual performance. This would allow for identifying DKE trends in unexamined contexts, and further confirm ecological validity and reliability of the DKE. This leads to the second prediction that, similar to previous research, bottom performers would overestimate performance while top performers underestimated. Bottom performers were also expected to display higher levels of estimation error.

Finally, this study aimed to determine the extent of DKE moderation by methodological factors such as time point for when performance estimations were made (i.e., pre v post), the reference group used for comparison (student v athletes), and individual current or prior sporting experience. This would also help account for variable findings in prior studies, confirm DKE robustness, and provide insight toward underlying etiologies. Therefore, the final prediction of this study was that increased experience and pre-task estimations would increase participant estimation accuracy, with limited differentiation between reference groups.

3.1 Methods

3.1.1 Participants

Participants were $N = 56$ (female = 22; M age = 23.5, $SD = 3.43$) undergraduate and post-graduate student volunteers at The University of Sydney, who reported a range of prior and current sporting experience (i.e., participation in 0-5 sports for between 0 – 24 years - $M = 5.12$ years; current cumulative participation ranged from 0 – 10 hrs/week - $M = 2.65$ hrs/week). Inclusion criteria specified that participants were between 18-30 years old, were healthy, and had no injury or reason limiting or preventing them from physical activity engagement. Full disclosure of study purposes was not provided so as not to affect participant responses.

3.1.2 Procedure

Following the University's ethical approval (Appendix C), participants refrained from smoking, consuming caffeine, alcohol or any other artificial substances within 12 hours of participation. Participants were recruited via flyers posted throughout the University of Sydney, Faculty of Health Science Campus (Appendix D). Participants initially provided consent and completed the Physical Activity Readiness Questionnaire and You (Appendix E; CSEP, 2002), and items regarding their current and past physical activity involvement (Appendix F). Individually, participants then participated in two physical tasks, namely a 60m Sprint and Vertical Jump in a randomized order, and were advised that other students were participating. These tasks were chosen as they represent normative and standard physical tasks (e.g., Kistler, Walsh, Horn, & Cox, 2010; Rösch et al., 2000) which most, if not all, participants would have experienced either in school physical education, sport participation, fitness assessment, or as part of general physical activity behavior. Participants could receive ongoing performance feedback during the task (e.g., visual, perceived effort),

though neither specific quantitative performance feedback was provided, nor did they see or observe other study participants.

3.1.2.1 60m Sprint

Participants were shown and advised that three consecutive sprint attempts were required on a concrete track (approx. 75m long) with a 2 minute recovery period between attempts. A 10-minute warm-up including jogging and a practice sprint was implemented. A Freelap Track & Field timing system (Freelap Australia, 2012) recorded sprint time, with a Touch transmitter positioned at the start, and a Junior transmitter positioned at 30m and 60m. Participants wore a Freelap Stopwatch on their waist, displaying exact sprint time for the researcher only.

3.1.2.2 Vertical Jump

Instruction and demonstration on how to jump using a counter-movement technique (see e.g., Bobbert, Gerritsen, Litjens, & Van Soest, 1996) was provided. Jump height was recorded using a Vertisonic apparatus (Lafayette Instrument, 2004), which uses sound propagation (i.e., echolocation) to determine Vertical Jump height. After a 10-minute warm-up and practice jump, three consecutive attempts were required with a 30 second recovery between attempts.

3.1.2.3 Task estimation

Immediately prior to respective actual task attempts, participants were asked two single item questions (Appendix G). Firstly, and for example, “*Compared to the average student at your University, where would you rank your performance time on a 60m sprint task?*” (aka - Estimated Student Rank Pre-Performance); and secondly, “*Compared to the average athlete your age (+/- 3yrs) outside your University, where would you rank your performance time on a 60m sprint task?*” (aka – Estimated Athlete Rank Pre-Performance). Each item was rated precisely on a 10cm horizontal line from 0-100 percentile ranks, with

descriptive anchors of 0 (*worst*), 50 (*average*) and 100 (*best or highest*) highlighted. An explanation of percentile rank was also provided. Immediately following task attempts, the same two questions were repeated (i.e., Estimated Student Rank Post-Performance; Estimated Athlete Rank Post-Performance) with questions presented in the same format and style as pre-performance. All participants completed both tasks and questions, with a ten-minute rest period between tasks. Three participants chose not to complete the 60m Sprint. Data were recorded using Appendix H. After participation, participants were emailed a participant debrief statement, providing participants with study specific information (Appendix I).

3.1.3 Data analysis

All data analyses were performed using SPSS (IBM Corp, 2012), and similar analysis procedures were completed for both tasks. Following Kruger and Dunning (1999), the mean of trial attempts was ranked according to actual performance percentile rank (1-100), with higher percentiles corresponding to better mean performance. Participants were then grouped according to actual performance quartile (i.e., 1 = lowest & 4 = highest Performance Quartile).

To verify that raw performance data had linear trends and differed according to Performance Quartile, a one-way ANOVA with trend-analysis was conducted. A threshold of 90% of the combined between groups sum of squares was set, and assumption tests identified that outliers were not evident. Normality and homogeneity of variance were not violated. Appropriate preliminary checks were conducted on the following analyses.

3.1.3.1 Overall estimation alignment

To determine overall correlations between Actual Performance Rank and estimations, and within each Performance Quartile according to this studies first prediction, Pearson's correlations were calculated. Correlations determine the linear relationship between two variables, and so would allow for the determination of participants overall capability for self-

assessment accuracy. The lower the correlations (i.e., closer to -1 than to 1) the more incapable individuals are at aligning estimates of performance with actual performance. Prior to calculating correlations, ratings for the four estimation items were combined into Estimation-Time and Reference-Group variables according to Table 2. This allowed for the isolation of the effect of Estimation-Time or Reference-Group on correlations. Correlation sizes were reported according to criteria set by Cohen (1988).

Table 2: Equations of combined Estimation-Time and Reference-Group variables.

Estimation -Time	Combined Estimated Rank Pre-Performance	=	Estimated Student Rank Pre- Performance	+	Estimated Athlete Rank Pre- Performance	/2
	Combined Estimated Rank Post-Performance	=	Estimated Student Rank Post- Performance	+	Estimated Athlete Rank Post- Performance	/2
Reference- Group	Combined Estimated Rank Student-Reference	=	Estimated Student Rank Pre- Performance	+	Estimated Student Rank Post- Performance	/2
	Combined Estimated Rank Athlete-Reference	=	Estimated Athlete Rank Pre- Performance	+	Estimated Athlete Rank Post- Performance	/2

Note. Averaging two of the four estimation items in four unique combinations created the four combined estimated variables, allowing for isolation of Estimation-Time and Reference-Group effects.

To determine the nature of slope characteristics for Actual Performance Rank and each of the combined estimated performance rank across Performance Quartiles, linear regressions were conducted and compared. Performance Quartile acted as the independent variable (IV) with Actual Performance Rank as the dependent (DV). This procedure was repeated with each of the four combined estimated performance ranks as DVs. Slope formulae were determined to predict each actual or estimated percentile rank using the straight line expression $y = ax + b$; where y = actual or estimated performance rank and x = Performance Quartile. Confidence intervals were calculated and compared for overlap, as this is considered the ideal method for reporting significance (Thompson, 2002). No overlap indicated a significant difference between slopes.

3.1.3.2 Estimation-performance difference

Continuing to test the first prediction, specific paired *t*-tests were conducted to determine the extent of estimation inaccuracy in each Performance Quartile, and show DKE existence. These paired Actual Performance Rank with each of the combined estimated ranks.

3.1.3.3 Estimation-performance error

To assess whether the magnitude and direction of estimation inaccuracy were related to performers in different Performance Quartiles and test the second prediction, independent *t*-tests were conducted. For these, four new variables of Estimation Error were created. For instance, Pre-Performance Estimation Error = (Combined Estimated Rank Pre-Performance – Actual Performance Rank); while Student-Reference Estimation Error = (Combined Estimated Rank Student-Reference – Actual Performance Rank). The other two were Post-Performance Estimation Error and Athlete-Reference Estimation Error.

3.1.3.4 Estimation moderation

Finally, to test the third prediction and determine whether original item estimates of performance (i.e., not combined) were significantly affected by Estimation-Time (i.e., pre v post task) or by Reference-Group (i.e., student v athlete) several additional analyses were performed. First, a 2 x 2 (Estimation-Time x Reference-Group) repeated measures ANOVA was performed using the four types of estimation rank as DVs (e.g., Estimated Student Rank Pre-Performance; Estimated Athlete Rank Post-Performance) and Performance Quartile as IV; this permitted between and within-group main contrasts. Then, multiple one-way ANOVA's with trend-analysis were conducted on each estimated rank permitting between and within-group interaction contrasts. To determine whether estimation differences existed between Estimation-Time and Reference-Group within each Performance Quartile, follow up paired *t*-tests were conducted using the combined estimated rank variables. Pairings of

Combined Estimated Rank Pre-Performance with Combined Estimated Rank Post-Performance, and Combined Estimated Rank Student-Reference with Combined Estimated Performance Athlete-Reference were used as DVs with Performance Quartile as IV.

3.1.3.5 Task experience

To continue testing the third prediction and determine whether prior experience affected actual Performance Quartile and Estimation Error, two sets of repeated independent *t*-tests were conducted. The first set assessed whether current athletic activity (i.e., either high v low) affected Actual Performance Rank and the four types of Estimation Error, while the second assessed the influence of previous sporting experience (i.e., either high or none). High current sporting activity was classified as > 4 cumulative hrs/week of sporting activity, while low was classified as 0-4 hrs/week. High previous sporting experience was considered having played ≥ 3 sports for ≥ 5 years, with low having played ≤ 1 sport for < 5 years. These groups reflected the extreme ranges of experience whilst maintaining sufficiently high numbers in comparison groups (i.e., $n \geq 10$).

3.2 Results

3.2.1 60m Sprint

The initial one-way ANOVA with trend-analysis identified a significant linear trend component across Performance Quartile on raw performance data ($F(1,50) = 68.99; p < .001, \eta^2 = .74$). Mean and SDs for Actual Performance Rank and each combined estimated performance rank according to Performance Quartile are shown in Table 3 and Figure 6.

3.2.1.1 Overall estimation alignment

Small to medium overall, and small to large isolated quartile specific correlations existed between Actual Performance Rank and the four types of estimated performance rank (see Table 4). Q1 and Q4 performers were the least calibrated, while Q2 and Q3 were more

accurate. Regression formulas for Actual Performance Rank and each of the combined estimated performance ranks are displayed in Table 5. The slope for Actual Performance Rank was significantly different to slopes for combined estimated percentile ranks.

3.2.1.2 Estimation-performance difference

Paired *t*-tests between Actual Performance Rank and the four combined estimated performance ranks according to Performance Quartile (see Table 3) highlighted significant differences for each combined estimated rank in Q1 and Q4 ($M = 31$ percentile points). No differences were apparent for Q2 and Q3 ($M = 7$ percentile points).

3.2.1.3 Estimation-performance error

In terms of direction of Estimation Error between Q1 and Q4 performers, significant differences existed for all four Combined Estimates - Pre-Performance ($t(24) = 7.63, p < .001$); Post-Performance ($t(24) = 8.40, p < .001$); Student-Reference ($t(24) = 7.72, p < .001$), and Athlete-Reference ($t(24) = 8.00, p < .001$). Q1 performers significantly overestimated ($M = 31$ percentile points), while Q4 performers significantly underestimated ($M = -30$ percentile points) performance relative to their Actual Performance Rank (see Figure 6 for illustration).

The magnitude of Estimation Error (i.e., non-directional) was not significant for any of the four Estimation Error variables, suggesting Q1 and Q4 performers were similarly error prone albeit in different directions. Estimation Errors between Q2 and Q3 only showed significant direction ($t(24) = 2.09, p = .047$) and magnitude ($t(25) = 2.45, p = .020$) differences for Combined Estimated Rank Student-Reference. The other three remained non-significant. That said, Q2 and Q3 error remained less than Q1 or Q4.

3.2.1.4 Estimation moderation

Results examining whether estimates were affected by Estimation-Time (i.e., pre v post task) or by Reference-Group (i.e., student v athlete), identified a significant linear trend

Table 3: Actual and combined estimated percentile ranks for 60m Sprint and Vertical Jump according to performance quartile.

	Performance Quartile		Actual Performance Rank	Combined Estimated Rank Pre-Performance	Combined Estimated Rank Post-Performance	Combined Estimated Rank Student-Reference	Combined Estimated Rank Athlete-Reference
60m Sprint	1	M	13.21	39.46*	50.23*	46.54*	43.15*
	(n=13)	(SD)	(7.34)	(20.44)	(21.58)	(20.25)	(21.72)
	2	M	37.74	39.08	40.92	45.54	34.46
	(n=13)	(SD)	(7.35)	(25.29)	(29.69)	(25.54)	(31.06)
	3	M	63.20	52.71	52.96	56.71	48.96
	(n=14)	(SD)	(7.89)	(13.91)	(15.33)	(15.10)	(18.27)
Vertical Jump	4	M	88.68	58.08*	58.50*	61.85*	54.73*
	(n=13)	(SD)	(7.34)	(14.25)	(12.57)	(16.69)	(12.13)
	1	M	13.39	48.32*	48.00*	51.86*	44.46*
	(n=14)	(SD)	(7.47)	(18.93)	(18.07)	(17.85)	(20.94)
	2	M	38.39	45.39	47.79	48.64	44.54
	(n=14)	(SD)	(7.47)	(24.90)	(21.86)	(22.09)	(25.62)
Vertical Jump	3	M	63.39	59.86	58.57	61.29	57.07
	(n=14)	(SD)	(7.46)	(16.71)	(15.49)	(14.52)	(19.12)
	4	M	88.39	59.14*	59.29*	63.89*	54.54*
	(n=14)	(SD)	(7.47)	(13.61)	(15.41)	(13.57)	(18.78)

Notes: * $p \leq .001$, * = Paired t -test identifying significant difference to Actual Performance Rank in the same Performance Quartile.

for the main between-group contrast of estimated performance ranks according to Performance Quartile ($F(1,49) = 5.05, p = .029, \eta^2 = .09$); that is, estimations generally increased along with Performance Quartile. Within-group contrasts collapsed across Performance Quartile, identified significance for Estimation-Time ($F(1,49) = 4.82, p = .033, \eta^2 = .09$) and Reference-Group ($F(1,49) = 10.56, p = .002, \eta^2 = .17$) but with no interaction ($F(1,49) = 1.32, p = .25$). This suggests estimates were independently influenced by time point of estimation and reference group irrespective of Performance Quartile.

Figure 6: Actual and estimated performance ranks for Sprint performance according to Performance Quartile.

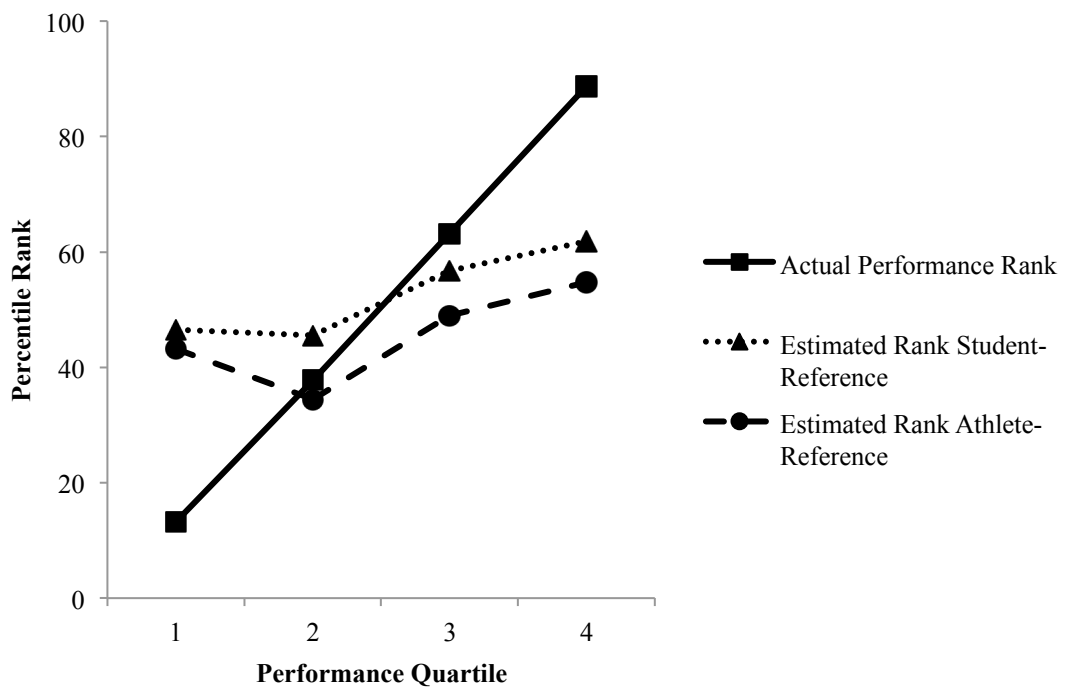
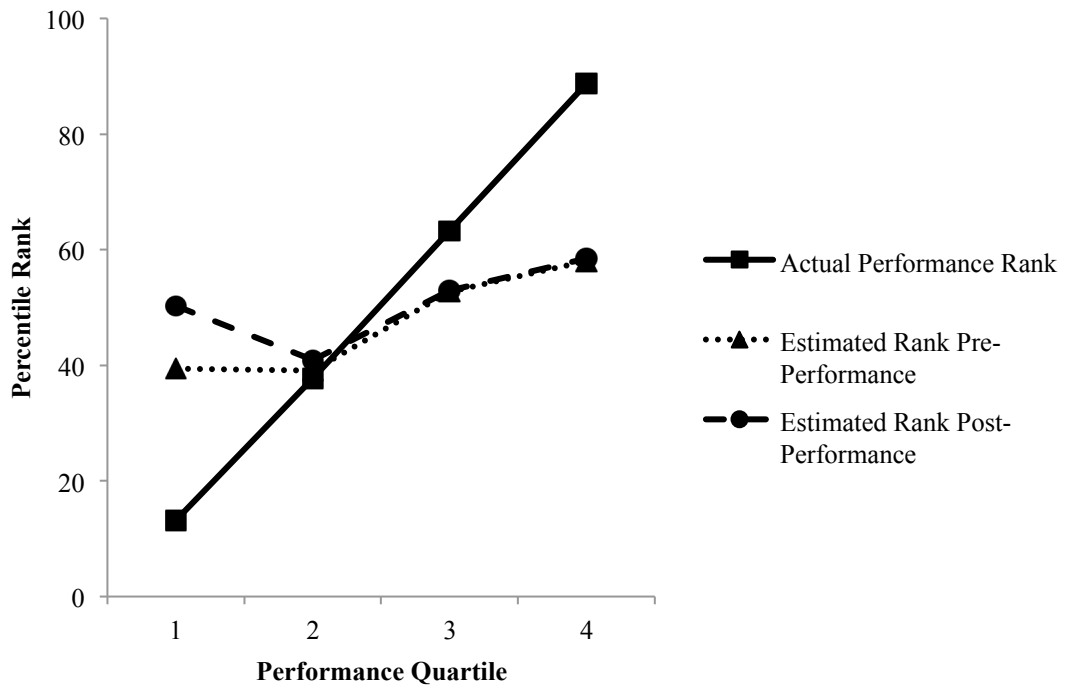


Table 4: Overall and quartile specific correlation coefficients between Actual Performance Rank and each combined estimated performance rank for the 60 m Sprint and Vertical Jump.

Task	Percentile Rank	Overall	Q1	Q2	Q3	Q4
60m Sprint	Combined Estimated Rank Pre-Performance	.43**	-.04	.64*	.21	.08
	Combined Estimated Rank Post-Performance	.25	-.42	.73*	.26	.17
	Combined Estimated Rank Student Reference	.38*	-.16	.68*	.42	.23
	Combined Estimated Rank Athlete Reference	.28*	-.30	.66*	.03	-.04
Vertical Jump	Combined Estimated Rank Pre-Performance	.25	-.11	-.20	-.04	.26
	Combined Estimated Rank Post-Performance	.26	.08	-.38	.01	.24
	Combined Estimated Rank Student Reference	.30*	.05	-.13	-.07	.17
	Combined Estimated Rank Athlete Reference	.20	-.07	-.40	-.02	.26

Notes: * $p \leq .05$, ** $p \leq .001$, * = significant correlation between Actual performance Rank and corresponding combined estimated performance rank in the same Performance Quartile.

When examining changes in estimation with Performance Quartile included, significant between-group interaction contrasts were evident for Estimation-Time ($F(1,49) = 5.73, p = .021, \eta^2 = .10$) but not Reference-Group ($F(1,49) = 0.15, p = .700$). The three-way interaction was not significant ($F(1,49) = 0.01, p = .940$). Thus, changes in estimates between pre and post time points were Performance Quartile dependent, while changes due to referent group were not. Paired t -tests between Estimation-Time and Reference-Group within each Performance Quartile did not isolate any specific mean differences (see Table 6). Descriptive indications ($p = .029$; Bonferroni adjustment, $p = .025$) showed Q1 performers as having the largest difference in estimates at pre v post time points (i.e., increase by 10 percentile points), and did the least estimation adjustment when reference groups were changed (i.e., student – athlete = -3 percentile points).

Table 5: Regression slopes for actual and combined estimated performance percentile ranks for 60m Sprint and Vertical Jump according to Performance Quartile.

Task	Percentile Rank	Slope	Height	95% CI [LL, UL]	<i>t</i>	<i>F</i>	<i>p</i>
60m Sprint	Actual Performance Rank	25.19	-12.26	[23.36, 27.02]	27.67	765.68	<.001
	Combined Estimated Rank Pre-Performance	6.96	29.96	[2.28, 11.64]	2.99	8.93	.004
	Combined Estimated Rank Post-Performance	3.69	41.44	[-1.51, 8.89]	1.42	2.03	.161
	Combined Estimated Rank Student Reference	5.72	38.39	[0.87, 10.57]	2.37	5.61	.002
	Combined Estimated Rank Athlete Reference	4.93	33.02	[-0.54, 10.40]	1.81	3.28	.076
Vertical Jump	Actual Performance Rank	25.00	-11.61	[23.24, 26.76]	28.54	814.61	<.001
	Combined Estimated Rank Pre-Performance	4.69	41.45	[0.14, 9.25]	2.07	4.27	.044
	Combined Estimated Rank Post-Performance	4.46	42.25	[0.22, 8.71]	2.11	4.44	.040
	Combined Estimated Rank Student Reference	4.88	44.23	[0.72, 9.03]	2.35	5.54	.022
	Combined Estimated Rank Athlete Reference	4.28	39.46	[-0.79, 9.68]	1.69	2.86	.096

Notes: CI = confidence interval for slope, LL = lower limit, UL = upper limit, *p* values signify whether the slope is significantly greater than zero.

Table 6: Comparison of estimated performance ranks types (pre v post; student v athletes) in Sprint and Vertical Jump according to performance quartile.

Performance Quartile	60m Sprint		Vertical Jump	
	Pre v Post	Student v Athlete	Pre v Post	Student v Athlete
1	-2.19	0.10	na	0.13
2	-0.89	2.01	na	1.59
3	0.47	1.74	na	0.82
4	0.04	1.93	na	2.99*

Note: * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$, * = paired *t*-test identifying significant difference between comparison estimates in the same Performance Quartile.

3.2.1.5 Task experience

Previous or current sporting experience was not found to affect Actual Performance Rank or types of Estimation Error (i.e., $p > .05$), suggesting that while Performance Quartile held specific relationships with Estimation Error (i.e., Q1 & Q4 more error prone), experience did not affect (i.e., for better or worse) estimation rank or error.

3.2.2 Vertical Jump

One-way ANOVA with trend-analysis identified a significant linear trend component across Performance Quartile on raw performance data ($F(1,52) = 449.27; p < .001, \eta^2 = .90$). Mean and SDs for Actual Performance Rank and combined estimated performance ranks in Performance Quartiles are shown in Table 3 and Figure 7.

3.2.2.1 Overall estimation alignment

Correlations between Actual Performance Rank and the four types of estimated performance rank (see Table 4) were non-significant (with exception of Combined Estimated Rank Student Reference). Only a weak correlation for Combined Estimated Student Reference was evident. This suggests that estimation was not well calibrated with performance in Vertical Jump. Regression formulas for Actual Performance Rank and each

combined estimated performance ranks (see Table 5) showed that the slope for Actual Performance Rank as significantly different to the slopes for estimated percentile ranks.

3.2.2.2 Estimation-performance difference

Paired *t*-tests between Actual Performance Rank and the four combined estimated performance ranks according to Performance Quartile are shown in Table 3. Significant differences between Actual Performance Rank and each of the combined estimated ranks were apparent for Q1 and Q4 ($M = 32$ percentile points), with no significant difference for Q2 and Q3 ($M = 6$ percentile points).

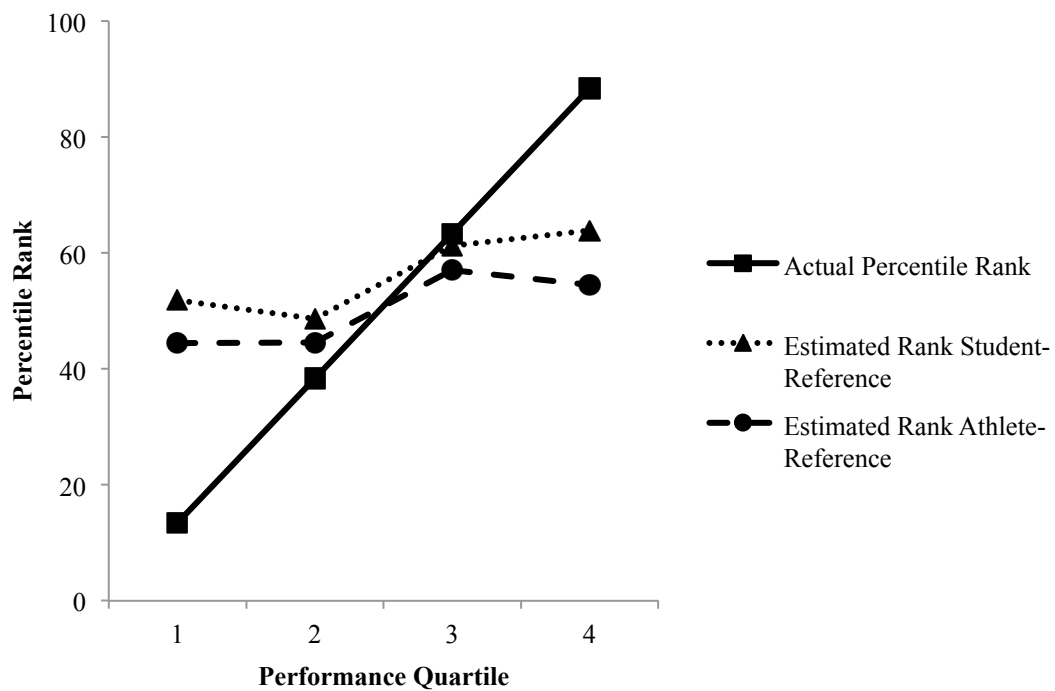
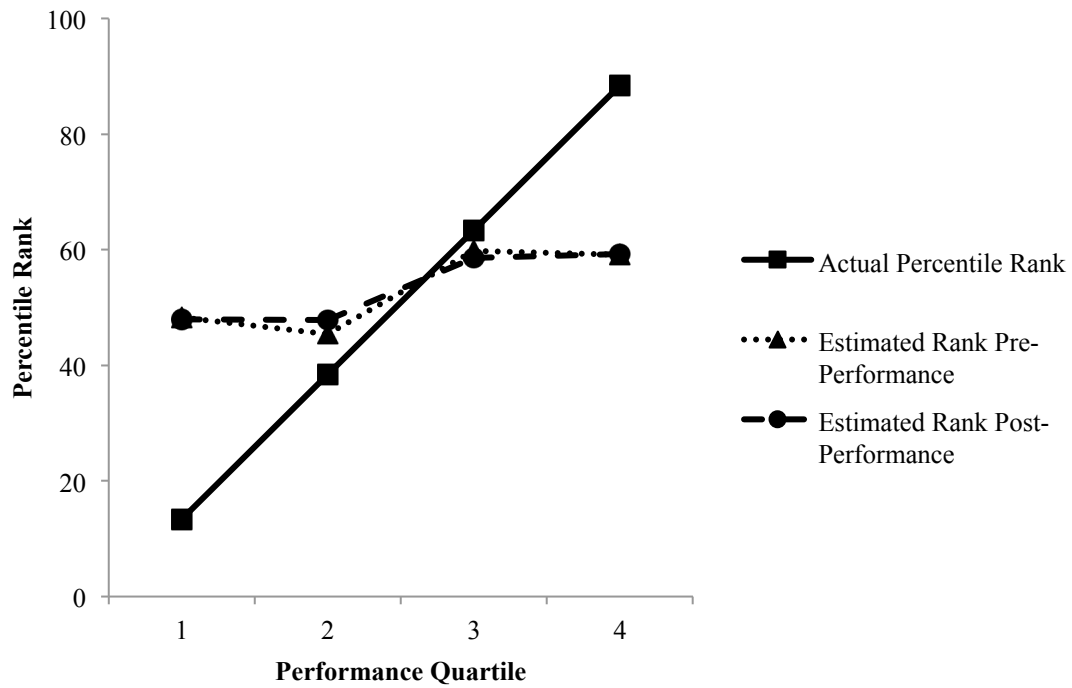
3.2.2.3 Estimation-performance error

In terms of Estimation Error direction between Q1 and Q4 performers, independent *t*-tests identified significant differences for all Combined Estimate Ranks, namely - Pre-Performance ($t(26) = 9.54, p < .001$); Post-Performance ($t(26) = 9.76, p < .001$); Student-Reference ($t(26) = 9.90, p < .001$); and Athlete-Reference ($t(26) = 8.33, p < .001$). Q1 performers significantly overestimated ($M = 35$ percentile points) while Q4 performers significantly underestimated ($M = 29$ percentile points) relative to their Actual Performance Rank (see Figure 7). The magnitude of Estimation Error (i.e., non-directional) was not significant for three of the Combined Estimation Error variables, with Student-Reference the exception ($t(26) = 2.24, p = .034$). Thus, Q1 and Q4 performers were similarly error prone in terms of magnitude. Direction and magnitude of estimation error was also similar between Q2 and Q3 performers.

3.2.2.4 Estimation moderation

Determining whether estimates were affected by Estimation-Time (i.e., pre v post task) or by Reference-Group (i.e., student v athlete), a one-way ANOVA was conducted. A significant linear trend for the main between-group contrast for estimated performance ranks

Figure 7: Actual and estimated performance ranks for Vertical Jump performance according to Performance Quartile.



according to Performance Quartile was found ($F(1,52) = 4.53, p = .038, \eta^2 = .08$). This indicated that estimations generally increased with Performance Quartile. Within-group

contrasts collapsed across Performance Quartile then identified significance for Reference-Group ($F(1,52) = 10.24, p = .002, \eta^2 = .16$), but not for Estimation-Time ($F(1,52) = 0.45, p = .832$) or the interaction ($F(1,52) = 1.19, p = .280$). This suggests that only reference group influenced estimates in Vertical Jump. When Performance Quartile was included, non-significant between-group interaction contrasts were evident for Reference-Group ($F(1,52) = 0.12, p = .730$) and the three-way interaction ($F(1,52) = 0.43, p = .510$), suggesting that estimate changes were not Performance Quartile dependent.

3.2.2.5 Task experience

Previous or current sporting experience did not affect Actual Performance Rank or any of the Estimation Error types in Vertical Jump based on categories utilized (i.e., $p > .05$). So while Performance Quartile had specific relationships with Estimation Error, prior or current experience did not influence Actual Performance Rank or Estimation Error.

3.3 Discussion

To date, research determining DKE pervasiveness beyond cognitive domains and into other external and ecologically valid contexts such as sport and physical activity has been limited. Further, limitations and methodological inconsistencies between and within studies, such as accounting for task experience, the timing of self-estimation, and the reference group used for comparison, have cast doubt on DKE's validity and its underlying mechanisms. In addressing these points, findings from examination of more familiar, or at least more commonly experienced sport and physical activity contexts, confirmed DKE's prevalence. A general limited capability for participants to accurately align estimations with actual performance was evident. Similar to prior studies (e.g., Mikkelsen et al., 2005), only small to moderate overall correlations existed between actual and estimated performance in the Sprint ($r = .28 - .43; p \leq .05$) and Vertical Jump ($r = .30; p \leq .05$). This indicates that in both

tasks, individuals were unable to strongly align estimations with actual performance. However, individuals displayed higher accuracy compared to previous research in cognitive domains (e.g., Gross & Latham, 2012). Particular types of performers (i.e., Q1 & Q4 performers) accounted for these figures. That said, making accurate estimates relative to others in the Vertical Jump appeared to be more challenging compared to the Sprint. This may be explained by the reduction in feedback from ongoing task interaction (i.e., decreased performance duration), and/or the lower tangibility in visibly detecting performance differences in Vertical Jump.

Adhering and supporting prior DKE studies (e.g., Ehrlinger et al., 2008; Kruger & Dunning, 1999), typical DKE trends were identified in both Sprint and Vertical Jump tasks, providing validity and reliability in physical and ecologically valid contexts. Differences in mean estimated performance and actual performance were clearly evident for Q1 and Q4 performers (see Figure 1 & 2). Q1's consistently overestimated relative to performance rank (i.e., $M = 31$ & 35 percentile points for Sprint & Vertical Jump respectively), while Q4's consistently underestimated (i.e., $M = -30$ & -29 percentile points). However, unlike prior studies (e.g., Ehrlinger et al., 2008; Kruger & Dunning, 1999), while the direction of estimation error between Q1's and Q4's remained different, the magnitude of estimation error was similar at both ends of the performance range.

Perhaps due to the combination of high self-estimation and comparatively low competency, and the association of Q1's being 'over-confident', previous studies have emphasized the negative consequences and implications of Q1 overestimation. Indeed, Q1's have indeed been linked (to list a few) with having lowered perceived learning needs (Albanese et al., 2006; Violato & Lockyer, 2006), an increased likelihood of causing physical harm via personal negligence (Burson et al., 2006), drowning (Petras et al., 2012), and medical in-competence (Haun et al., 2000). Yet without ignoring these concerns, present

findings also suggest that Q4 error should not necessarily be overlooked. For instance, the inability of top-performers to accurately self-evaluate may be as equally debilitating in terms of motivation, as is metacognitive incapability for Q1 performers. For example, Mattern et al. (2010) showed how Q4 performers in higher education perceived their grades as being (incorrectly) lower compared to others, and which subsequently affected their persistence and graduation rates. So, if Q4 estimation error and understanding is not realigned with actual performance, it could lead to decreased learning investment (Cury et al., 1997) and less interest in potentially rewarding situations (Ehrlinger & Dunning, 2003).

In highlighting and addressing limitations and methodological inconsistencies in prior DKE studies, present findings identify that task experience (whether current or previous) did not associate with Performance Quartile, nor reduce any types of Estimation Error. This goes against popular assumption, yet is unsurprising as prior studies consistently indicate how simple task engagement or experience is not equated with better performance or skill (Ericsson, Krampe, & Tesch-Römer, 1993), nor does it relate to a reduction in perception or performance errors. The irrelevance of task experience in DKEs can also be explained by differences in metacognitive skill (i.e., the ability to think about one's thinking - Flavell, 1979; Kruger & Dunning, 1999). Individuals with lower metacognitive skill (i.e., Q1's) are associated with the failure to evaluate and identify factors leading to performance errors, and are less able to act and adjust due to a lack of domain knowledge and self-insight. Therefore, regardless of repeated experience they are unable to identify, interpret, and act upon available and relevant feedback. Instead, and based on present findings, their perceptions may be informed more by internal feelings and emotions (e.g., perceived exertion and affect). In contrast, individuals with higher metacognitive skills (i.e., more likely to be Q4's) may use a combination of ongoing external information (e.g., feedback, reference group comparison),

internal knowledge (e.g., explicit, procedural), and self-monitoring to guide their more cautious estimations.

A methodological inconsistency in prior DKE studies is the time-point of estimation (i.e., pre v post), which was found to significantly affect estimation in the Sprint. Specifically, post-task estimation actually increased estimation error (i.e., $M = 2.99$), suggesting that in certain contexts - and possibly more so for Q1 performers (see Figure 1) - post-task estimations may be confounded by other factors. Such influential factors are difficult to pinpoint, but estimate inflation could relate to task effort and investment, and/or feelings of affect (e.g., satisfaction) gained from task participation and completion. Certainly these aspects did not affect pre-post estimates of higher quartile performers, and again may reflect better metacognitive regulation in such individuals (Kruger & Dunning, 1999). The potential influence of chronic and possibly skewed ability perceptions has previously been highlighted (Critcher & Dunning, 2009; Pazicni & Bauer, 2014), but these should hypothetically affect both pre and post conditions, as opposed to post-task estimation alone. Nevertheless, the finding holds methodological implications for onward studies, as post-task estimation may encourage artificial inflation due to other influences in the lower performing quartiles.

The manipulation of reference comparison (i.e., student v athlete) also helped address DKE methodological concerns. Specifically, whether participant estimates were affected by such changes, whether estimates were intentionally adjusted given question item content, and carefully considered (i.e., reliable). Findings showed that athlete referenced estimates were consistently lower compared to student comparisons (i.e., Sprint - $M = -7.34$; Vertical Jump - $M = -7.80$), demonstrating that referent group characteristics affected estimation ratings, and that participants considered the reference comparison group in changing estimates. Only Q1's in the Sprint task did little estimate adjustment (see Figure 1), potentially suggesting that they

were still more influenced by immediate self and task related perceptions, and less by re-evaluation of self-standing relative to changes in social situation.

Due to the vast and consistent support for the metacognitive mechanism of the DKE as detailed in Chapter 2, this study sought to examine additional DKE factors. While this study indirectly supports the metacognitive mechanisms of the DKE, it was unable to directly investigate the relationship between individual metacognition and self-assessment accuracy. The results obtained in this study remain applicable for individual self-assessment however, regardless of the validity of metacognitive mechanisms. Additionally, this study refrained from analysing Q2 and Q3 performers. While the DKE can be best viewed through the extremes of Q1 and Q4 performers, this restriction may have led to the loss of additional insights into self-assessment accuracy.

Chapter 3 - 4 Bridging Statement

Chapter 2 identified contexts in which the DKE had yet to be identified, as well as methodological inconsistencies between DKE studies. Chapter 3 detailed the existence of the DKE in familiar physical contexts (i.e., Sprint and Vertical Jump), and assessed the effect of varying task experience, and how manipulation of timing of estimation (i.e., pre v post task), and reference groups (i.e., student v athlete) affected self-assessment accuracy. Upon completion of this study other questions remained and/or arose, such as whether self-assessment (in)accuracy and DKE's were more evident in less familiar tasks (i.e., with no previous task experience), whether task-efficacy could be considered as an influential participant variable, and whether manipulating task difficulty (i.e., high v low), along with the available inherent ongoing performance feedback (i.e., high v low) would also influence performance estimates.

Chapter 4 aimed to address and answer these questions by examining and identifying DKE prevalence and self-assessment (in)accuracy in two unique and unfamiliar cognitive tasks (i.e., Stroop test and The Tower of Hanoi). These tasks allowed for the removal of any potential influence of prior task experience or exposure, helping to isolate the influence of initial perceived task efficacy (possibly a proxy of chronic efficacy), due to their unfamiliarity. They also allowed for removal of the potential influence of difficulty by being either very easy or difficult to complete, as well as for investigation of the influence of task specific ongoing performance feedback on self-estimation accuracy. Similar to the study in Chapter 3, the timing of estimations (i.e., pre v post task), and the reference group (i.e., student v athlete) were again manipulated to assess estimation sensitivity across performance quartiles. This study was therefore designed to further develop our understanding of the individuals and the particular contexts where DKEs would be most prevalent, and where vulnerability to detrimental DKE outcomes may lie. It also aimed to further understand the

cognitive processes involved in (in)accurate self-assessments by analysing the changes in performance perceptions according to the methodological variations described; as well as help isolate who, where, how, and when prevention and intervention strategies might be effectively targeted.

Chapter 4 – The Dunning-Kruger Effect in Cognitive Performance Contexts

Abstract

Background: The inability of individuals to accurately align performance estimates with actual performance is reflected in the Dunning-Kruger effect (DKE; Kruger & Dunning, 1999). **Objectives:** This study sought to further DKE understanding by assessing whether estimation-performance alignment errors occur in two unfamiliar cognitive tasks (i.e., the Stroop and Tower of Hanoi) where task difficulty and feedback were manipulated, and whether chronic self-views (Ehrlinger & Dunning, 2003) could act as a DKE mechanism. The time point of estimation (i.e., pre v post) and the reference group (student v athlete) were also examined as potential DKE moderators. **Methods:** Based on procedures similar to prior DKE studies (i.e., Chapter 3), participant's actual performance was compared to their estimated performance, while determining the impact of the various potential moderators. **Results:** Findings illustrated poor general alignment of estimation-performance capability (e.g., $r = .21$; $r = -.12$). DKEs were clearly evident with Q1 performers significantly overestimating performance (i.e., $M = 58$ & 20 percentile points for Stroop & Tower of Hanoi respectively), with Q4 performers significantly underestimating ($M = -28$ & -41). Importantly, task efficacy was found to predict the degree of overestimation and underestimation, suggesting that chronic self-views can partially explain DKEs in task with no familiarity and experience. The DKE was moderated by the timing of the estimation, which interacted with task characteristics, as performance estimates increased after the Stroop ($M = 5$ points), but decreased after the Tower of Hanoi ($M = -13$ points). **Conclusions:** Together, findings again highlight the pervasiveness of the DKE. Participant, task, and methodological factors interacted to moderate the relative size of DKE's. Importantly, besides task specific and metacognitive knowledge, chronic and pre-existing self-views were supported as a mechanism. Discernible in-task feedback also appeared as a potential strategy for realigning estimation with performance.

Keywords: Perceived Performance, Actual Performance, DKE, Stroop, Tower of Hanoi

4.0 Introduction

Referring to a type of cognitive bias, the Dunning-Kruger Effect (DKE; Dunning, 2011; Schlösser et al., 2013) describes how self-assessment error, reflecting the difference between estimated and actual performance, occurs differentially as a function of task capability and domain specific knowledge. DKEs are commonly reflected by the tendency for bottom quartile (Q1) performers to consistently overestimate their performance when compared to others, while in contrast top (Q4) performers underestimate their performance on any given task. These trends have been identified in numerous domains and task contexts, including university students' logical reasoning ability (Ehrlinger et al., 2008), specialist physicians' clinical practice (Violato & Lockyer, 2006), and salesmen's ability to sell (Jaramillo et al., 2003), suggesting that DKEs are pervasive.

The signature DKE trend however, appears to be more variable than static. As identified in a recent meta-analysis (i.e., Chapter 2), estimation (in)accuracies vary according to: the task context (i.e., cognitive v physical; Kruger & Dunning, 1999; Mikkelsen et al., 2005); the time for when performance estimates are made (i.e., pre v post-performance; Ryvkin et al., 2012; Sheldon et al., 2014); and against whom the estimations are made (i.e., peers v average; Albanese et al., 2006; Ames & Kammrath, 2004). Additional data suggest that the availability of tangible performance feedback (Ferraro, 2010; Ryvkin et al., 2012), the degree of task difficulty (Burson et al., 2006), and prior task experience (Kruger & Dunning, 1999) may also influence estimation-performance alignment. As identified in Chapter 2, the moderation of these inaccuracies may also be task capability dependent. That is, the ability to utilise and act upon available information during tasks, and hence adjust estimations, may relate to whether individuals are among the lowest 25% (Q1) or top 25% (Q4) of performers. For example, Q4 performers may be beneficially affected by the timing of performance estimation, correctively adjusting performance estimates post-task following

task experience and feedback, whereas the experience for Q1's may serve only to increase their error, suggesting the operation of many interacting factors.

The sources of DKE (in)accuracies have been linked previously with a lack of task-specific and metacognitive skill in Q1 performers (Kruger & Dunning, 1999), and a lack of comparative information in Q4 performers (i.e., false-consensus effect; Ross et al., 1977). Though scarcely examined, alternative mechanisms have also been proposed and could help explain the variations in estimation (in)accuracy. Performance estimations for instance, may rely on previously constructed perceptions regarding performance capability (i.e., chronic self-views; Ehrlinger & Dunning, 2003), which require a 'minimum threshold' of knowledge or experience (Kruger & Dunning, 1999). Chronic self-views could lead performers to assess their task capability based on previous beliefs, rather than on actual task experience and familiarity. Optimistic self-views thus may influence psycho-social characteristics like task efficacy, influence pre-task estimation and explain overestimation in Q1 performers (Battistelli et al., 2009; Ehrlinger et al., 2008). By comparison, cautionary or pessimistic self-views can be reflected by modesty and lower efficacy, and may explain Q4 underestimation (Ehrlinger et al., 2008). If task completion and/or feedback then substantiates or reinforces these views, this could account for post-task inflation of both Q1 overestimation and Q4 underestimation.

Related more to Q1 performers, heightened self-estimates (i.e., overestimation) have also been hypothesized to result from egocentric weighting (Battistelli et al., 2009; Kruger, 1999). This implies that when individuals estimate their performance, they place more 'weight' on their own perceived capability than on the possible performance of others (Kruger, 1999), which could then lead to insufficient adjustment of estimations according to reference group changes. This study indirectly examines this mechanism by asking whether

performers in different performance quartiles modify their performance estimates when the reference group used for comparison is manipulated.

To investigate DKEs, DKE moderation, chronic self-views, and egocentric weighting mechanisms along with their relationships with estimation-performance error, it is necessary to expose and capture these aspects in representative task contexts. This exposes participants with no prior experience to highly unfamiliar tasks of varying difficulty and offers a feasible experimental and comparative approach to assessing the influence of chronic self-views. Also, asking participants to compare performance estimates against various comparison groups (e.g., student, athlete) can help determine whether egocentric weighting influences estimation error and whether it is related to performance capability (i.e., Q1-Q4). Such experimental manipulations will help provide a better understanding of both the subtleties and intricacy of the DKE phenomenon, and help establish the underlying mechanisms.

In the present study, two independent cognitive tasks were utilized. The first was the Stroop test, a test commonly used to assess executive cognitive functioning, such as the ability to inhibit a proponent response due to the conflict between linguistic and perceptual components of the stimulus presented in the task (Epp, Dobson, Dozois, & Frewen, 2012; MacLeod, 1991; Phillips, Bull, Adams, & Fraser, 2002; Pothos & Tapper, 2010). Although not previously used in DKE terms, the Stroop is generally unfamiliar, easy to complete, and could help expose pre-conceived self-views due to having no prior exposure. The task can be administered so as to provide limited feedback available during the task, have relatively simple standard instructions and requirements for completion, and with minimal expected performance differences between healthy individuals of a similar age (Van der Elst, Van Boxtel, Van Breukelen, & Jolles, 2006). The task can also be utilized to assess whether estimations are more or less adjusted by particular performers (i.e., Q1's v Q4's) by asking standard item questions pre and post-task.

By contrast, the Tower of Hanoi puzzle is a cognitive task that has been used to assess problem-solving capability, planning, working memory, solution strategies, and self-monitoring ability (Goel & Grafman, 1995; Kotovsky, Hayes, & Simon, 1985; Welsh, 1991). Whilst having the appearance of being a relatively simple task, it is actually difficult to complete during initial trials, thus hypothetically it could help expose chronic self-views and task efficacy in pre-task estimation. The task however does permit concurrent and tangible performance feedback, as success (or lack of it) can easily be observed and interpreted, therefore whether feedback can correct or mitigate estimation-performance error can also be examined. While these task contexts themselves can be compared, question items can also assess whether the timing of estimation (pre v post), and reference anchoring (student v athlete) moderate estimation errors; and likewise infer whether egocentric weighting processes appears to occur.

The purpose of this study was to determine DKE trends, as well as the degree and direction of estimation error (i.e., the difference between actual and estimated performance) within and across actual performance quartiles in two novel, unfamiliar (and one difficult), cognitive task contexts. Second, the study attempted to highlight the key interactions between specific task and methodological factors, and how they might moderate estimation-performance error. This included manipulation for when performance estimates were requested, and whom reference comparisons were made. Finally, the study sought to determine whether particular participant factors such as psycho-social characteristics, chronic self-views, and egocentric mechanisms could be exposed and be related to task over/underestimation, accounting for DKEs.

Therefore, this study made three predictions: Prediction 1 states that participants will display typical DKE trends, in that they will be unable to accurately self-assess performance. Prediction 2 states that bottom performing participants will significantly overestimate their

performance, top performers significantly underestimate their performance, and that top performers will be more accurate. Finally, prediction 3 states that participants will be more accurate prior to performance, will not differentiate between reference groups, and will become more confident following performance in an easy task.

4.1 Methods

4.1.1 Participants

Following University ethical approval (Appendix C), participants were $N = 56$ (female = 22; M age = 23.5, $SD = 3.43$) undergraduate and graduate student volunteers at The University of Sydney; and were recruited via flyers posted throughout the Faculty of Health Science Campus (Appendix D). Inclusion criteria specified that participants were aged between 18-30 years, were healthy, and not presently taking medication for any illness or condition. Full disclosure of study purposes was not provided so as not to affect participant responses in their estimations or performance (Appendix E). All participants indicated they were both unfamiliar and had no previous experience in completing the Stroop or Tower of Hanoi tasks.

4.1.2 Procedure

In the 12 hours prior to participation, participants were asked to refrain from smoking, consuming caffeine or alcohol, or any other stimulating or artificial substances. Individually, participants provided informed consent, and then completed two cognitive tasks in a randomized order within a private lab space. Participants were advised that other students of similar age were participating in the study.

4.1.2.1 Stroop

To conduct the Stroop test, a standard 15.5 inch (39.4cm) computer screen and purpose built latency timing device with accompanying software program (RL-Timer: Steel

& Eisenhuth, 2012) was used. The latency device had six choice keys arranged in a semi-circular pattern around a 'home key'. The three keys to the right were red, green, and blue in color, and were the responses for a right-handed individual. The left hand response keys were similarly arranged for a left-handed individual. Only one set of response keys was displayed dependent on participant handedness.

Requiring an average of 10 minutes, the Stroop test requested 60 responses to sequential word displays. Between displays, participants were able to rest for up to 30 seconds if desired. Participants were instructed to press and hold the 'home key', which after a delay of between 1-3 seconds, triggered displays of one of three words (i.e., RED, GREEN, BLUE) shown in one of the three colours. Of the total, 30 words were presented as congruent (i.e., spelling & colour matched), and 30 were incongruent. Upon word display, participants had to move their index finger as quickly as possible from the 'home key' to the corresponding coloured response key. The latency device determined reaction time (i.e., stimulus word display – pressure release of 'home key'), and movement time (i.e., pressure release of 'home key' – pressure applied to response key). The average total response time (i.e., reaction + movement time) across all 60 responses were recorded for data analysis.

4.1.2.2 Tower of Hanoi

A 7-ring version of the Tower of Hanoi (similar to Goel & Grafman, 1995) was used for this study due to its initial perceived simplicity, actual degree of task difficulty, high levels of ongoing performance feedback, and potentially large variation in relative task success. Participants were shown the task and informed of the aims (i.e., moving rings from the left to right peg), and rules (i.e., moving one ring at a time; cannot place a larger ring on top of a smaller ring) for puzzle completion. A time limit of 10 minutes was provided for task completion, and participants were advised that their performance would be compared to other participating students using the following scoring criterion. Scoring criteria included (in

hierarchical order) completion time (e.g., Goel & Grafman, 1995), the number of rings successfully stacked (e.g., Welsh, Satterlee-Cartmell, & Stine, 1999), and total moves made (e.g., Goldberg, Saint-Cyr, & Weinberger, 1990). Participants were percentile ranked using an amalgamation score. If performances had identical values on a given criteria (e.g., two participants completed in identical time or didn't complete the task), subsequent criteria were applied to determine rank (i.e., rings stacked successfully or number of moves undertaken). The amalgamation and percentile score were recorded for data-analysis.

4.1.2.3 Task estimation

Immediately prior to respective task attempts, participants were asked three single item questions (Appendix G). Firstly, "*Compared to the average student at your university, where would you rank your performance time in the Tower of Hanoi task?*" (Estimated Student Rank Pre-Performance); secondly, "*Compared to the average student at your university, how confident are you in being able complete the task close/near to your personal best?*" (Pre-Performance Efficacy); and thirdly, "*Compared to the average athlete your age (+/- 3yrs) outside your university, where would you rank your performance time in the Tower of Hanoi task?*" (Estimated Athlete Rank Pre-Performance). Each item was rated precisely on a 10cm horizontal line from 0-100 percentile ranks, with descriptive anchors of 0 (*worst*), 50 (*average*), and 100 (*best or highest*) percentile ranks. An explanation for percentile rank was provided. The wordings of items were also adapted according to the task. Immediately after completing the Stroop and Tower of Hanoi, participants were again asked the same three items (i.e., Estimated Student Rank Post-Performance, Post-Performance Efficacy, Estimated Athlete Rank Post-Performance). Each item was presented using the same format and style as pre-task estimation items. Data was recorded using Appendix H. After participation, participants were emailed a participant debrief statement, providing participants with study specific information (Appendix I).

4.1.3 Data analysis

All data analyses were performed using SPSS v.21 (IBM Corp, 2012). For both tasks similar data analysis steps were conducted. Initially, to determine whether original performance data had a significant trend according to performance quartile, a one-way ANOVA with trend-analysis was conducted. A threshold of 90% of the combined between-groups sum of squares was set, and tests identified that outliers were not evident, and that normality and homogeneity of variance were not violated. Preliminary checks were also conducted on the subsequent analyses. Similar to Kruger and Dunning (1999), raw performance data was then percentile ranked (1-100), with higher percentiles corresponding to better performance. Participants were then grouped according to Performance Quartile (i.e., 1 = lowest & 4 = highest quartile). Data for one student in the Tower of Hanoi was incomplete and not included.

4.1.3.1 Overall estimation alignment

To determine the overall capability of individuals to align performance estimations with actual performance and test prediction 1, Pearson's correlations between Actual Performance Rank and each estimated performance rank were conducted. As correlations determine the linear relationship between two variables, the higher the correlation (i.e., closer to 1 than to -1) the more accurate individuals are at assessing their performance. For this to be done, the four estimated-performance rankings were combined according to Estimation-Time and Reference-Group variables as seen in Table 2. Effect sizes of correlations were reported according to Cohen (1988).

4.1.3.2 Estimation-performance error

To assess whether the magnitude and direction of estimation inaccuracy were related to performers in different Performance Quartiles and test prediction 2, independent *t*-tests were conducted. For these, four new variables of Estimation Error were created. These were

Pre-Performance Estimation Error = (Combined Estimated Rank Pre-Performance – Actual Performance Rank); and Student-Reference Estimation Error = (Combined Estimated Rank Student-Reference – Actual Performance Rank). The other two variables were Post-Performance Estimation Error and Athlete-Reference Estimation Error. These four types then acted as DVs with Performance Quartile as IV. To further determine whether Q1 and Q4 Estimation Error were significant, specific paired *t*-tests were conducted between Actual Performance Rank and each of the combined estimated ranks within both Performance Quartiles.

4.1.3.3 Estimation moderation

To test prediction 3, and to determine whether the original item estimations (i.e., not combined) were significantly affected by Estimation-Time (i.e., pre v post task) and hence chronic-self views or feedback, or by Reference-Group (i.e., student v athlete) and hence egocentrism, additional analyses were performed with the following steps. First, a 2 x 2 (Estimation-Time x Reference-Group) repeated measures ANOVA was performed using the four types of estimation rank acting as DV's (e.g., Estimated Student Rank Pre-Performance; Estimated Athlete Rank Post-Performance) and Performance Quartile as IV. This permitted both between and within-group main contrasts. Then, multiple one-way ANOVA's with trend-analysis on each estimated rank permitted computation of both between and within-group interaction contrasts. To determine whether estimation differences existed between Estimation-Time and Reference-Group within each Performance Quartile, follow up paired *t*-tests were conducted using the combined estimated rank variables. Pairings of Combined Estimated Rank Pre-Performance with Combined Estimated Rank Post-Performance, and Combined Estimated Rank Student-Reference with Combined Estimated Performance Athlete-Reference were used as DVs with Performance Quartile as IV.

4.1.3.4 Task efficacy

Continuing to test prediction 3, and to determine whether Q1 performers displayed higher task- efficacy than Q4 performers both prior to and following task completion, independent-sample *t*-tests were conducted. To determine whether cognitive task characteristics significantly impacted on participant pre and post task efficacy, paired-sample *t*-tests were used. To determine the relationship between both pre and post-task efficacy and estimation error linear regressions were conducted. Pre-task efficacy acted as the IV and estimation Error as the DV, and this was repeated for post-task efficacy. Slope formulas were calculated using: $y = ax + b$; where $y =$ Estimation Error, and $x =$ Task Efficacy.

4.2 Results

4.2.1 Stroop

The initial one-way ANOVA with trend-analysis identified a significant linear trend component across Performance Quartile on average response times ($F(1,55) = 173.26; p < .001$). As expected however, the range of actual performance was small (i.e., 0.34 ms) suggesting minimal performance differences between participants. When converted to Actual Performance Rank, the Mean and SDs and each combined estimated performance rank according to Performance Quartile are shown in Table 7 and Figure 8. Data shows that regardless of participant unfamiliarity with the task, participants consistently provided above average estimations of performance, irrespective of actual performance rank. Q1 performers not only rated their performance as above average, they consistently estimated higher than Q4 performers.

4.2.1.1 Overall estimation alignment

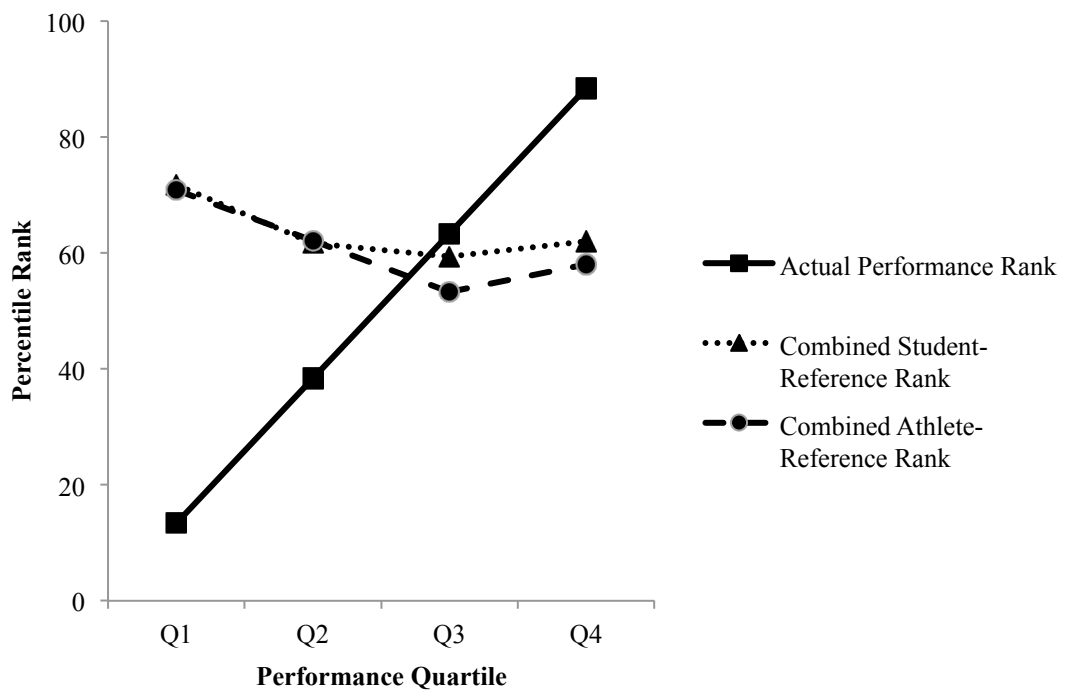
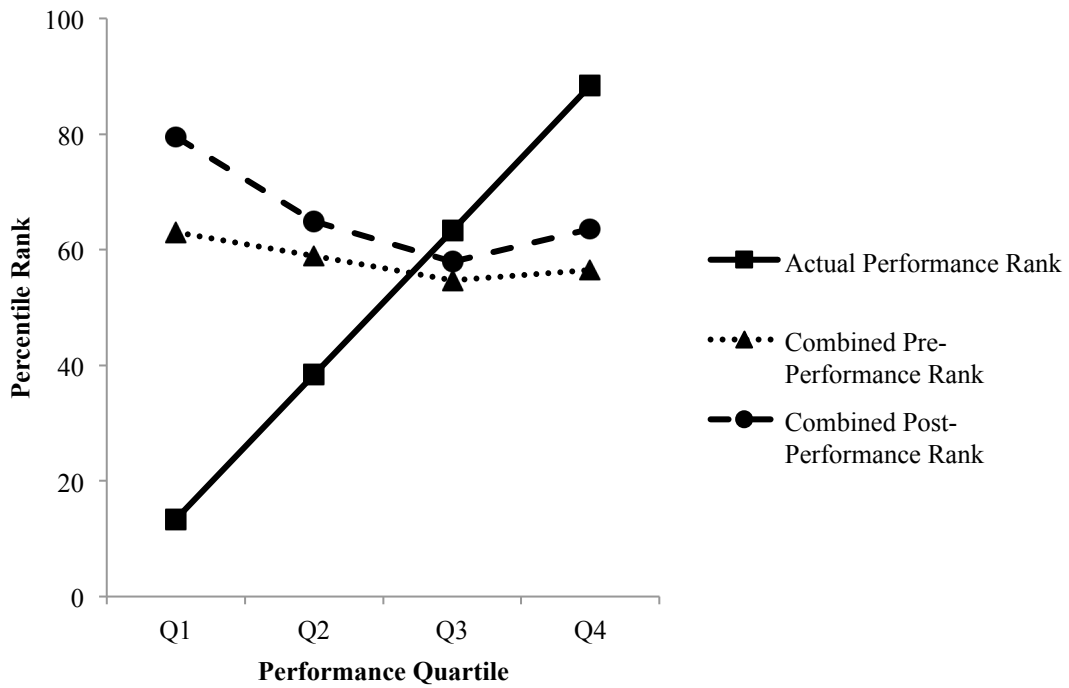
Significant overall correlations existed between Actual Performance Rank and Combined Estimated Rank Post-Performance ($r = -.37; p = .005$), Combined Estimated Rank

Table 7: Actual and estimated percentile ranks for Stroop Test and Tower of Hanoi according to performance quartile.

	Performance Quartile		Actual Performance Rank	Combined Estimated Rank Pre-Performance	Combined Estimated Rank Post-Performance	Combined Estimated Rank Student-Reference	Combined Estimated Rank Athlete-Reference	Pre-Performance Efficacy	Post-Performance Efficacy
Stroop	1	M	13.39	63.07**	79.57**	71.75**	70.89**	68.79	82.93
	(N=14)	(SD)	(7.47)	(14.93)	(17.25)	(10.71)	(-14.95)	(14.26)	(14.78)
	2	M	38.39	58.96**	64.86**	61.68**	62.14**	65.57	67.86
	(N=14)	(SD)	(7.47)	(14.53)	(13.76)	(12.08)	(-13.96)	(16.85)	(14.10)
	3	M	63.39	54.71*	57.93	59.39	53.25*	66.57	62.14
	(N=14)	(SD)	(7.47)	(9.59)	(12.95)	(10.74)	(-8.25)	(18.48)	(14.03)
	4	M	88.39	56.50**	63.54**	62.00**	58.04**	68.57	70.64
	(N=14)	(SD)	(7.47)	(17.78)	(12.68)	(15.19)	(-15.3)	(20.50)	(13.83)
Tower of Hanoi	1	M	12.72	44.85**	20.77	31.69*	33.93*	48.00	16.54
	(N=13)	(SD)	(7.07)	(24.03)	(20.67)	(20.44)	(-22.88)	(27.49)	(20.79)
	2	M	37.27	44.39	25.04	35.03	34.43	51.79	21.71
	(N=14)	(SD)	(7.60)	(21.02)	(14.75)	(15.96)	(-16.64)	(26.27)	(12.51)
	3	M	62.73	40.50*	21.50**	29.79**	32.21**	44.79	17.64
	(N=14)	(SD)	(7.60)	(19.33)	(18.76)	(17.54)	(-20.02)	(24.77)	(15.28)
	4	M	88.18	55.46**	38.89**	44.96**	49.39**	63.93	32.43
	(N=14)	(SD)	(7.60)	(22.22)	(26.60)	(23.23)	(-22.98)	(22.72)	(25.06)

Note: * $p \leq .0125$, ** $p \leq .001$, * = paired t -test identifying significant difference to Actual Performance Rank in the same Performance Quartile.

Figure 8: Actual and estimated performance ranks for Stroop performance according to performance quartile.



Student-Reference ($r = -.27; p = .048$), and Combined Estimated Rank Athlete-Reference ($r = .35; p = .007$); though not Combined Estimated Rank Pre-Performance ($r = -.18; p = .181$).

This indicates a general small-moderate capability of individuals to accurately align estimations with actual performance.

4.2.1.2 Estimation-performance error

In terms of the direction of Estimation Error between Q1 and Q4 performers, significant differences existed for all four Combined Estimates: Pre-Performance ($t(26) = 12.24, p < .001$); Post-Performance ($t(21.12) = 14.40, p < .001$); Student-Reference ($t(26) = 14.94, p < .001$), and Athlete-Reference ($t(26) = 14.18, p < .001$). Q1 performers significantly overestimated ($M = 58$ percentile points), while Q4 performers significantly underestimated ($M = -28$ percentile points) relative to their Actual Performance Rank (see Figure 8 for illustration). The magnitude of Estimation Error (i.e., non-directional) was also significantly different for each of the four Estimation Error variables, Pre-Performance ($t(26) = 2.69, p = .013$); Post-Performance ($t(19.47) = 6.64, p < .001$); Student-Reference ($t(26) = 5.69, p < .001$), and Athlete-Reference ($t(26) = 4.40, p < .001$). Q1 performers were significantly more prone to error compared to Q4 performers, indicating Q1 performer estimation error was both larger and in the opposite direction compared to Q4 error. Estimation Errors between Q2 and Q3 also showed significant direction and magnitude differences for all combined estimation ranks. That said, Q2 and Q3 displayed less error compared to Q1 or Q4. Paired t -tests between Actual Performance Rank and each of the four combined estimated performance ranks according to Performance Quartile (see Table 7) highlighted significant Estimation Error for both Q1 and Q4 performers.

4.2.1.3 Estimation moderation

Examining whether estimates were affected by Estimation-Time (i.e., pre v post task) or by Reference-Group (i.e., student v athlete), identified a significant linear trend for the main between-group contrast of estimated performance ranks according to Performance Quartile ($F(1,52) = 7.27, p = .009, \eta^2 = .11$). That is, estimations generally decreased across

performance quartile. Within-group contrasts collapsed across Performance Quartile, identified significance for Estimation-Time ($F(1,52) = 16.36, p = .001, \eta^2 = .22$) and Reference-Group ($F(1,52) = 6.13, p = .017, \eta^2 = .10$), but with no significant interaction ($F(1,52) = 1.16, p = .286$). This suggests that estimates were independently influenced by time point of estimation and reference group irrespective of Performance Quartile.

When examining changes in estimation with Performance Quartile included, no significant between-group interaction contrasts were evident for Estimation-Time ($F(1,52) = 2.97, p = .091$) or Reference-Group ($F(1,52) = 2.82, p = .099$). The three-way interaction was also not significant ($F(1,52) = 2.15, p = .149$). Thus, changes in estimates between pre and post time points and reference group were not dependent on Performance Quartile. Paired *t*-tests between Estimation-Time and Reference-Group variables within each Performance Quartile (see Table 8) isolated significant mean differences between pre and post-performance estimates for Q1 and Q4 performers ($p \leq .012$), as well as significant differences between student and athlete reference groups for Q3 ($p = .006$). This indicates that actual performance and task feedback (i.e., post-performance estimates) increased initial perceptions based on chronic self-views (i.e., pre-performance estimates).

Table 8: Comparison of estimated performance ranks types (pre v post; student v athletes; efficacy) in Stroop Test and Tower of Hanoi according to performance quartile.

Performance Quartile	Stroop			Tower of Hanoi		
	Pre v Post	Student v Athlete	Pre v Post Efficacy	Pre v Post	Student v Athlete	Pre v Post Efficacy
1	-2.97*	0.39	-2.71*	6.95***	na	5.82***
2	-1.58	-0.20	-0.42	4.10***	na	4.75***
3	-0.85	3.24**	0.73	5.52***	na	3.19**
4	-2.91*	1.95	-0.54	3.16**	na	4.32***

Note: * $p \leq .025$, ** $p \leq .01$, *** $p \leq .001$, values correspond with calculated *t*-values; * = paired *t*-test identifying significant difference between each of the four combined estimated performance ranks in the same Performance Quartile. Significance adjusted according to Bonferroni correction. na= not attempted, due to lack of evidence for reference group mediation in ANOVA result.

4.2.1.4 Task efficacy

Independent sample *t*-tests indicated Q1 performers were equally as confident as Q4 performers prior to the Stroop ($t(26) = 0.03, p = .975$), yet perceived that they were significantly more efficacious after completion ($t(26) = 2.27, p = .032$), even though they had performed the worst. Paired sample *t*-tests indicated significant pre and post efficacy changes in Q1 performers only (see Table 8). Linear regressions between Pre-Task Efficacy and Pre-Task Estimation Error ($y = .67(x) - 36.09; p = .011; R^2 = .11$), and between Post-Task Efficacy and Post-Task Estimation Error ($y = 1.44(x) - 85.64; p < .001; R^2 = .37$) were significant, indicating that increased efficacy was predictive of performance overestimation, while decreased efficacy was predictive of performance underestimation (see Figure 9).

4.2.2 Tower of Hanoi

Initial one-way ANOVA with trend-analysis identified a significant linear trend component across Performance Quartile on raw performance data ($F(1,54) = 908.37; p < .001$). As expected, the range of raw performance scores indicated greater performance variability (i.e., taking less than 9 minutes to complete the task in 254 moves – to taking all allotted time and only moving 4 rings in a total of 200 moves). Mean and SDs for Actual Performance Rank and each combined estimated performance rank according to Performance Quartile are shown in Table 7 and Figure 10. Data showed that participants provided consistently below average (i.e., 50%) estimations of performance rank, with the exception of Q4 Combined Estimated Rank Pre-Performance.

4.2.2.1 Overall estimation alignment

Significant overall correlations existed between Actual Performance Rank and Combined Estimated Rank Post-Performance ($r = .29; p = .031$) but not for Combined Estimated Rank Pre-Performance ($r = .12; p = .404$), Combined Estimated Rank Student-Reference ($r = .20; p = .137$), or Combined Estimated Rank Athlete-Reference ($r = .22; p =$

.106). This indicated a small-moderate capability to align estimation with actual performance that was only apparent after task completion.

Figure 9: Linear regression between Pre and Post-Performance Efficacy and Estimation

Error in Stroop task.

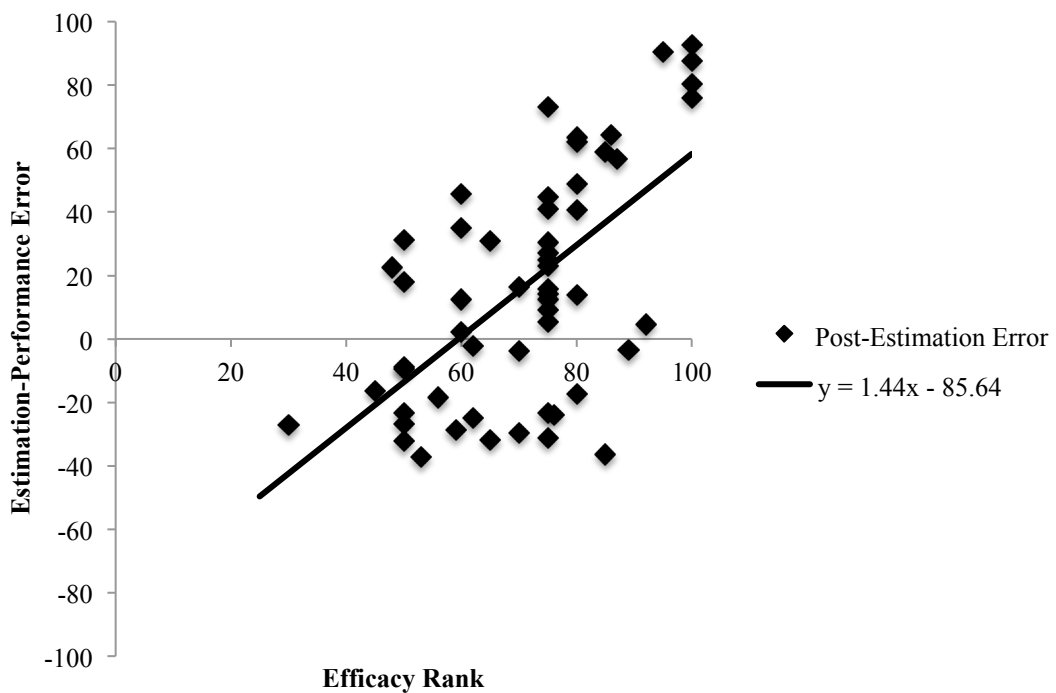
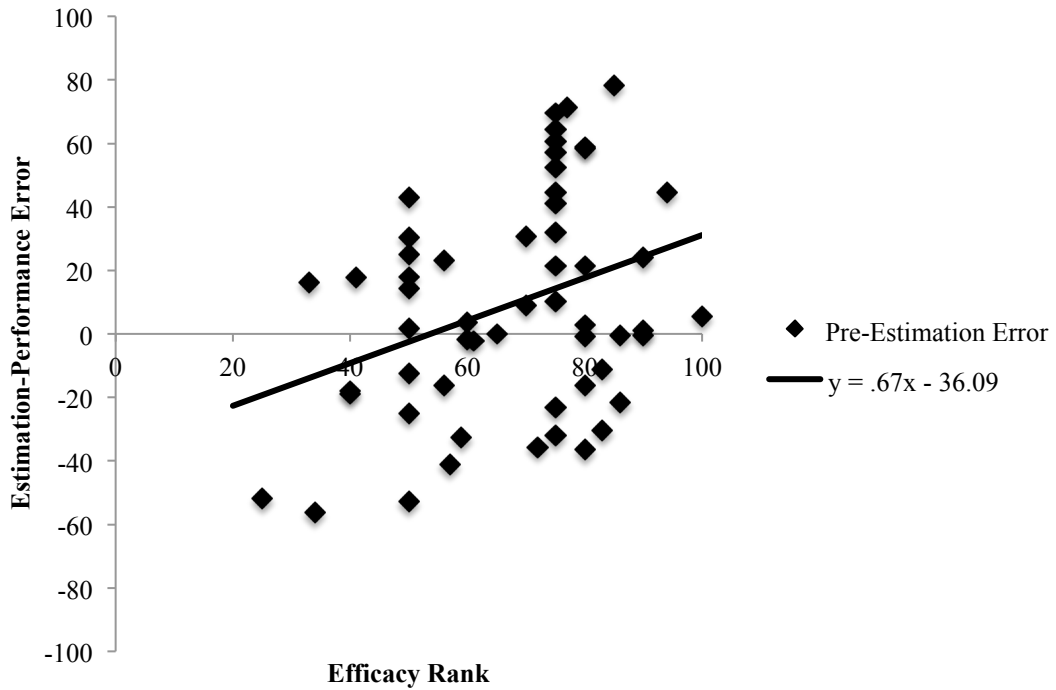
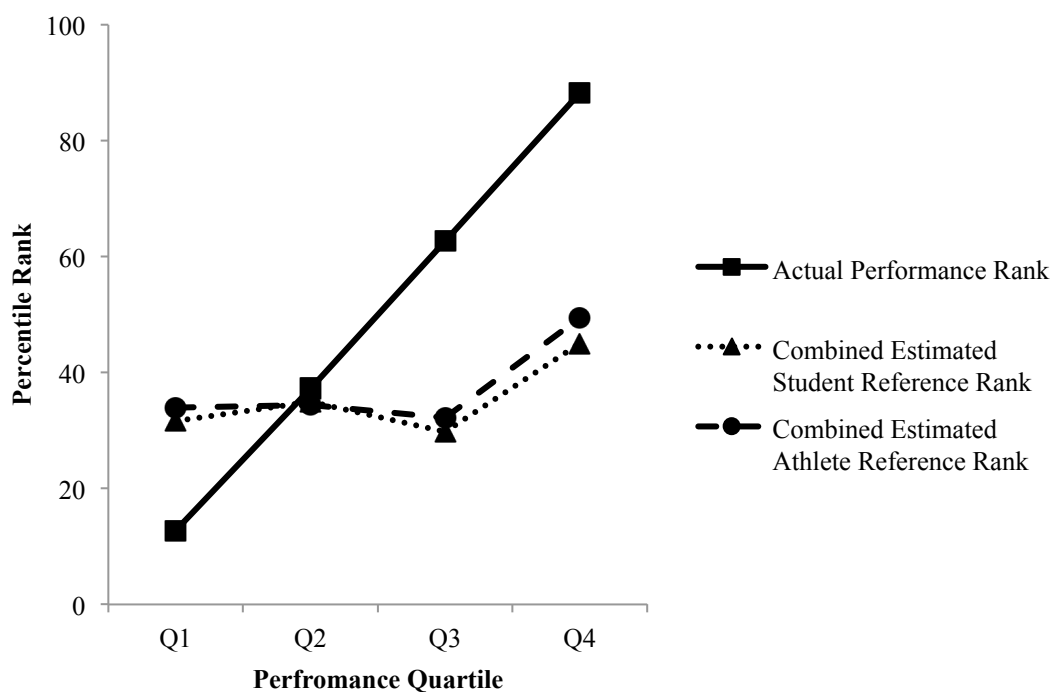
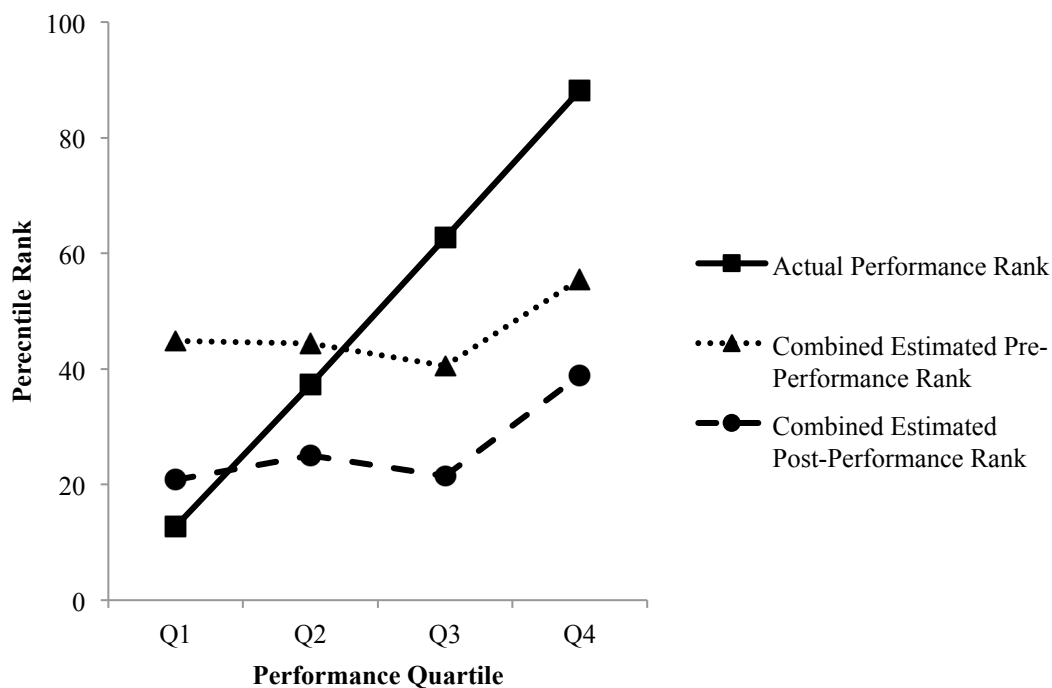


Figure 10: Actual and estimated performance ranks for Tower of Hanoi performance according to performance quartile.



4.2.2.2 Estimation-performance error

In terms of the direction of Estimation Error between Q1 and Q4 performers, significant differences existed for all four Combined Estimates - Pre-Performance ($t(25) = 7.29, p < .001$); Post-Performance ($t(25) = 6.60, p < .001$); Student-Reference ($t(25) = 7.71, p < .001$); and Athlete-Reference ($t(25) = 6.94, p < .001$). Q1 performers significantly overestimated ($M = 20$ percentile points), while Q4 performers significantly underestimated ($M = -82$ percentile points) relative to their Actual Performance Rank (see Figure 10 for illustration). The magnitude of Estimation Error (i.e., non-directional) was also significantly different for two Estimation Error variables: Post-Performance ($t(25) = -4.32, p < .001$) and Student-Reference ($t(25) = -3.15, p = .004$); but not Pre-Performance ($t(26) = 2.69, p = .013$) and Athlete-Reference ($t(26) = 4.40, p < .001$). Q4 performers were error prone to a greater magnitude and in the opposing direction of error, compared to Q1 performers. Estimation Errors between Q2 and Q3 showed similar significant direction and magnitude differences for all combined estimation ranks. That said, Q1 and Q2 displayed less error compared to Q3 or Q4. Paired t -tests between Actual Performance Rank and the four combined estimated performance ranks according to Performance Quartile (see Table 7) highlighted significant Estimation Error for both Q1 and Q4, with exception of the comparison between Actual Performance Rank and Combined Estimated Rank Post-Performance for Q1. Together, this suggests that performance estimations were significantly different than their actual performance, with Q1's becoming more accurate post-task.

4.2.2.3 Estimation moderation

Results examining whether estimates were affected by Estimation-Time (i.e., pre v post task) or by Reference-Group (i.e., student v athlete), did not identify any significant trend for the main between-group contrast of estimated performance ranks according to Performance Quartile ($F(1,54) = 2.74, p = .104$). That is, estimations generally remained

constant across Performance Quartile. Within-group contrasts collapsed across Performance Quartile, identified significance for Estimation-Time ($F(1,51) = 83.52, p \leq .001, \eta^2 = .61$) but not Reference-Group ($F(1,51) = 2.93, p = .093$) or the interaction ($F(1,51) = 0.41, p = .526$). This suggests that while estimates were not influenced by the manipulation of Reference Group, nor through any Performance Quartile interaction, they were independently influenced by time point of estimation (i.e., chronic self-views and performance feedback).

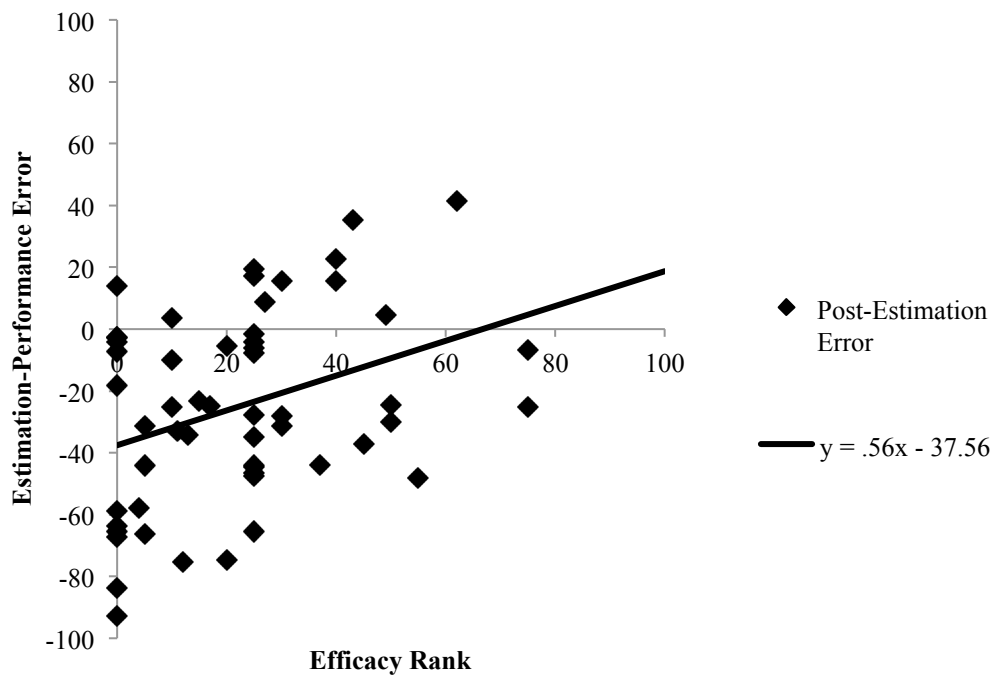
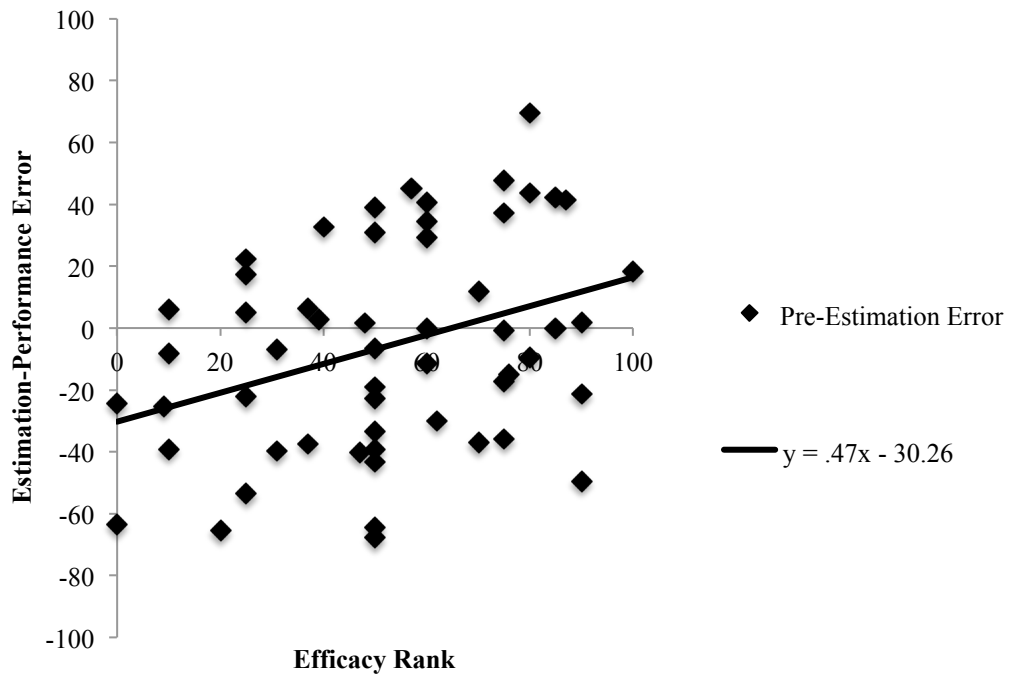
When examining changes in estimation with Performance Quartile included, no significant between-group interaction contrast was evident for Estimation-Time ($F(1,54) = 0.51, p = .246$), and the three-way interaction was not significant ($F(1,54) = 0.79, p = .377$). Thus, changes in estimates between pre and post time points were not Performance Quartile dependent. Paired t -tests between Estimation-Time within each Performance Quartile identified that pre-post task changes in estimates for each quartile ($p \leq .008$).

4.2.2.4 Task efficacy

Independent sample t -tests indicated Q1 performers were equally confident as Q4 performers both prior to task performance ($t(26) = -1.65, p = .112$) and after ($t(26) = -1.79, p = .086$). Paired t -tests also indicated significant pre-post efficacy changes in all performers (see Table 8). Linear regressions between Pre-Task Efficacy and Pre-Task Estimation Error ($y = 0.47(x) - 30.26; p = .009; R^2 = .12$), and between Post-Task Efficacy and Post-Task Estimation Error ($y = 0.56(x) - 37.56; p = .009; R^2 = .12$) identified that those with lower efficacy were predictive of performance underestimation, and how those with higher efficacy were predictive of overestimation (see Figure 11).

Figure 11: Linear regression between Pre and Post-Performance Efficacy and Estimation

Error in the Tower of Hanoi.



4.3 Discussion

Being set within two cognitive task contexts where participants had no previous experience or exposure, DKEs remained pervasive. This finding was represented in a poor/limited capability to align estimation with performance, and by the significant degree and direction of estimation error within and across actual performance quartiles. Specific participant, task characteristic features, and methodological factors moderated estimation-performance error. Finally, the study determined that participant chronic self-views (reflected by task efficacy) were likely exposed prior to the tasks, and were predictive of over and underestimation, thus helping to account for how DKEs occur. More of this is revealed in the task specific findings.

4.3.1 Stroop

Because it is easy to perform with minimal in-task feedback, all participants completed the Stroop within relatively small ranges of response times (i.e., 0.34 ms). When performance was percentile ranked, correlations with estimation were non-significant suggesting that the lack of experience and exposure made alignment challenging, arguably more so than in previous tasks examined (e.g., intelligence, Brim, 1954; information literacy, Gross & Latham, 2012; logic, Kruger & Dunning, 1999). Contrary to the notion of a ‘minimum threshold’ (Kruger & Dunning, 1999) where a minimum level of knowledge or experience is necessary for Q1 overestimation, a lack of task experience did not prevent an emergence of DKEs. For instance, pre-performance estimates were generally above average, with Q1 performers overestimating their performance by an average 50 percentile points, while Q4 performers underestimated their performance (i.e., -32 percentile points). These trends remained evident post-task (i.e., Q1 overestimation = 66 percentile points; Q4 underestimation = -25 percentile points), and so findings align to existing literature indicating

DKE pervasiveness across task contexts that provided minimal task related feedback (Ames & Kammrath, 2004; Ehrlinger et al., 2008; von Stumm, 2013).

Estimation timing moderated Stroop DKEs, with significant increases in post-task estimates for Q1 performers, as they became more erroneous in their overestimation (i.e., 17 percentile points). By contrast, task experience led Q4 performers to become more accurate, with reductions in their degree of underestimation (i.e., 7 percentile points). These findings may be due to rises in performance efficacy (and estimation) based on perceptions of performance success (Elzubeir & Rizk, 2000; Greifeneder et al., 2010). Findings may also align with suggestions from Burson et al. (2006) and Kruger (1999) in terms of egocentric weighting. As the Stroop may have been perceived as relatively easy - without feedback to suggest anything to the contrary - this may account for post-task estimation inflation and due to a lesser consideration for the performance of others. While there were no hypothetically expected differences to exist between student and athlete performances in the Stroop, this manipulation did affect overall estimation ranks with a general reduction in estimation - independent of quartile - when comparing to athletes. This adjustment did not occur for any specific quartile (e.g., Q1's), and so provides mixed indication of egocentric weighting (i.e., ignoring of reference group information) overall.

Findings related to task efficacy suggest that chronic self-views (Ehrlinger & Dunning, 2003) can account for DKEs. For instance prior to the Stroop, efficacy ratings were relatively optimistic regardless of having no task experience; though initial task instructions and observation may have positively influenced efficacy ratings. Regression uniquely identified that pre-task efficacy significantly predicted estimation error, with low levels of pre-task efficacy predicting underestimation (more likely in Q4's) and high pre-task efficacy predicting overestimation (more likely in Q1's). Without in-task feedback, efficacy (along with error) was also found to increase post-task and was likely related to initial perceptions of

task success (Feltz, 1988; Greifeneder et al., 2010) and chronic self-views (Ehrlinger & Dunning, 2003), as shown by the strengthening relationship between post-task efficacy and estimation error. The combination of no experience and metacognition (Kruger & Dunning, 1999) thus seems to have exposed chronic self-views (i.e., higher efficacy) - obtained from other partially or non-related experiences and tasks – affecting estimation error. Further, as feedback has been associated with increased metacognitive monitoring (Butler et al., 2008), the minimal degree of feedback provided during the Stroop may also explain the exacerbation of Q1 overestimation.

4.3.2 Tower of Hanoi

Set again as an unfamiliar task, this task contained relatively simple aims and instructions yet was difficult to complete during initial trial attempts (as reflected by large variations in performance). The Tower of Hanoi however, did evidently provide observable and tangible in-task feedback relating to performance progress (or lack of). Findings identified that pre-task correlations between estimation and actual performance rank were non-significant, with alignment improving post-task; indicating experience assisted individual estimation accuracy. The task did induce lower ‘average’ pre-task performance estimates when compared to the Stroop, supporting the idea of a worse-than-average bias and suggesting some possible initial uncertainty. Nonetheless, general DKE trends prevailed with Q1 performers significantly overestimating performance pre-task, while Q4 performers significantly underestimated performance. Across all participants, post-task estimates were generally reduced, with Q1 performers showing more accurate estimations, while Q4 continued to underestimate. This degree of Q1 accuracy is not typically seen in the DKE literature (Ehrlinger et al., 2008; Kruger & Dunning, 1999; Ryvkin et al., 2012), and may be partially due to the increased performance feedback.

DKEs in the Tower of Hanoi were moderated by estimation timing, but in this case (compared to the Stroop) significant reductions in post-task estimates for Q1 performers were apparent, while Q4's exacerbated their underestimation. Research indicates that as participants receive feedback indicating high task difficulty, and more specifically failure, then predictable decreases in performance estimation (and efficacy) occur (Greifeneder et al., 2010; Gutin et al., 2006). Some authors have suggested that such feedback increases vigilance and metacognitive monitoring (Butler et al., 2008) subsequently leading to revised and reduced estimation errors. Alternatively, as the Tower of Hanoi was viewed as difficult, individuals may have egocentrically lowered estimates based on personal performance relative to the known criteria as opposed to comparing to others per se (Burson et al., 2006; Kruger, 1999). The lack of overall and quartile specific adjustment of estimates according to reference group (i.e., student v athlete) suggests that egocentrism was influential in this task context.

Task efficacy findings also provided evidence that chronic self-views (Ehrlinger & Dunning, 2003) had a predictive role in DKE over/underestimation. Pre-task efficacy ratings were - by comparison to Stroop - more cautious and uncertain regarding their performance capability, and they may also have been influenced by initial pre-task instructions and observation. Nevertheless, pre-task efficacy significantly predicted estimation error, with low efficacy predicting underestimation and high efficacy predicting overestimation. With tangible in-task performance feedback available, efficacy (along with estimates) also generally decreased post-task, and a similar relationship between post-task efficacy and estimation error was evident. Again as participants had no prior experience or task knowledge, findings lend support to the hypothesis that pre and post-task perceptions were shaped by chronic self-views (Ehrlinger & Dunning, 2003). Importantly however, in-task

progress could be easily self-determined as the task was difficult to complete, and may therefore account for the observed mitigation of Q1 overestimation.

4.3.3 Implications and recommendations

Findings from this study highlight the potential for DKEs to be apparent in new or unfamiliar daily contexts, such as learning in education settings and performing novel tasks in the workplace. Chronic self-views, or beliefs about one's capability based on prior experiences elsewhere, may be important to consider prior to such task engagement, and in preventing the negative consequences from over and underestimation. Previously, Q1 overestimation has been associated with lower perceptions of training needs (Jaramillo et al., 2003), laboratory mistakes (Haun et al., 2000), physical injury (Burson et al., 2006), and unsafe behaviour (De Craen et al., 2011); whereas Q4 underestimation has been associated with lowered educational attainment (Mattern et al., 2010), reduced task involvement (Cury et al., 1997), and rejection of opportunities (Ehrlinger & Dunning, 2003). Thus, the avoidance and prevention of such outcomes would be beneficial.

Research into how DKE inaccuracies can be effectively prevented or mitigated (see Chapter 2) has been limited. Of the attempts, those targeting metacognitive skill have been the most successful (Kruger & Dunning, 1999), while those providing either forms of post-task feedback (e.g., Ferraro, 2010; Ryvkin et al., 2012), or monetary incentives (i.e., Ehrlinger et al., 2008) have not been effective. Future interventions that help develop metacognitive skills, or that help prevent the influence of overly positive or negative self-views will be valuable in avoiding DKE associated consequences. Strategies that individuals may utilize to increase metacognitive skill may include planning, goal setting, organization, and increased inherent performance feedback (Schraw, 1998; Zimmerman, 1990). In education and workplace settings, this may result in increased overall goal and daily performance alignment, increased lesson notes and record keeping, self-regulation of

continued learning, and constant evaluation of performance capability. In novel tasks, the realignment of estimates via clear discernible performance in-task feedback, and controlled task failure during initial exposure, may be highly valuable to Q1 overestimation; whilst an understanding of performance capability relative to others and awareness of hindering chronic self-views may be valuable for Q4 underestimation.

Due to the findings in Chapter 2, this study sought to examine factors that influenced self-assessment accuracy, rather than re-investigate the validity of metacognition as the DKE mechanism. The results obtained in relation to individual self-assessment research however, remain unique and provide significant contributions to understanding estimation accuracy. Additionally, this study refrained from analysing Q2 and Q3 performers, as self-assessment error is most apparent in the extreme performers. Including these groups in the analyses may have allowed for further understanding of cognitive bias and its effects, therefore future research into these individuals would be beneficial.

Chapter 5 – A Discussion of Dunning-Kruger Effect Studies in Physical and Cognitive

Contexts

Abstract

Analyzing the previous three Chapters and their investigation into the Dunning-Kruger Effect (DKE) allowed for an overall discussion regarding DKE mechanisms, moderators, implications, and future research directions. The DKE seems to be highly influenced by a combination of metacognitive skill, chronic-self views, and task efficacy. Increased performance feedback may increase metacognitive skill leading to increased estimation accuracy. Chronic self-views and task efficacy also influence estimation error, through may be mitigated by increased metacognitive skill. Originality of previous Chapters adds to research significance, while contributing to cognitive bias and DKE literature. Implications of inaccurate self-assessments exists for all individuals, with increased metacognitive skill seemingly the most effective strategy to increase accuracy. Limitations of the previous Chapters exist, directing additional avenues for future research

Keywords: Metacognition, Chronic Self-Views, Performance Feedback, Task Efficacy

5.0 Introduction

Since 1999, research into the cognitive bias known as the Dunning-Kruger effect (DKE; Dunning, 2011; Kruger & Dunning, 1999) has investigated the individual capability to accurately self-assess comparative performance. The consistent inaccuracies when attempting to align estimated and actual performance characteristic of the DKE (i.e., Q1 overestimation and Q4 underestimation) have significant real-world consequences for self-assessing individuals (e.g., Chapter 2; Mattern et al., 2010). However, methodological inconsistencies in DKE studies have led to gaps in the understanding of the mechanisms and cognitive processes that cause the DKE, diminishing the effectiveness of potential interventions designed to lessen DKE inaccuracies.

A recent series of studies (i.e., Chapters 2, 3, and 4) sought to mitigate the unknowns in DKE understanding, and were thus able to: summarize previous DKE research; identify potential DKE consequences, interventions, and mechanisms; determine the pervasiveness of the DKE paradigm; account for and confirm various methodological variations in DKE findings; highlight interactions between estimation (in)accuracy and various individual and task characteristics; and develop a unique understanding of how individuals self-assess performance capability. Therefore the purpose of this Chapter was to review and evaluate the contributions of Chapters 2, 3, and 4 to the current literature, entailing a detailed summary of each Chapter, and analysis of findings, originality, implications, limitations, and recommendations for future research.

5.1 Study Summary

Based on meta-analytical procedures applied to systematically identified studies, Chapter 2 investigated the alignment between estimated and actual performance characteristic of the DKE. When comparing relative to others, and across numerous task

contexts, findings identified only a small-moderate general capability to accurately match estimations with actual performance. Capability appears to be moderated by several study methodological factors, and varies according to task type, though was not affected by participant demographics (i.e., age, gender, or status). Significant pooled mean difference estimates, according to quartile categories of actual performance, indicate that (in)accuracy is associated with relative performance. Inter-individual differences, in acquired task-specific skill/knowledge, and metacognitive skill/awareness, were proposed to explain estimation-performance (mal)alignment, associated moderating influences (e.g., pre-post task estimation changes), and additional influences (e.g., psycho-social characteristics). While some negative consequences of mis-calibration were identified, determining further consequences within and across diverse tasks (e.g., driving), and contexts (e.g., education, medicine), would add to research impetus. The significance and relevance of estimation-performance capability could be substantial given the numerous domains, tasks, and specific situations where such (in)capabilities have been found, likely affecting individual learning, functionality, and health.

Chapter 3 investigated the existence of the DKE in unique, short, effortful, and familiar physical tasks (i.e., Sprint and Vertical Jump), and confirmed the DKEs prevalence. University student participants were somewhat (in)accurate when aligning their estimation and actual performance rank, with Q1 performers overestimating, and Q4 performers underestimating relative to actual task performance. Sporting experience had no influence on performance quartile or estimation error. Mal-alignment between estimation and performance was inflated when estimations were made following task performance (v pre-performance), while estimations were lowered when made according to an athlete reference group (v student reference group). The limited capability to accurately align task estimation with

actual performance at both ends of the performance spectrum was associated with original DKE mechanisms such as metacognitive skill.

Chapter 4 illustrates the pervasive nature of DKEs as a cognitive bias, identifying signature trends of Q1 overestimation and Q4 underestimation in two cognitive tasks, where participant task experience and knowledge was controlled. Findings identify how particular participant factors (e.g., task exposure, pre-existing experience in other activities and tasks), task factors (e.g., initial perceived difficulty, requirements content, inherent feedback availability), and methodological factors (e.g., time of point of measurement) interact to moderate DKE sizes. The present study is the first to produce evidence to suggest that chronic self-views - reflected by task-efficacy - can partially explain the direction of Q1 and Q4 estimation error, supporting an alternative DKE mechanism. Results did not consistently suggest that egocentric weighting occurred, rather indicating it to be task specific and situational in occurrence. Finally, changing the nature of in-task feedback affected the direction and size of post-task estimation error. In particular, low feedback in the Stroop compounded Q1 error, while self-determinable in-task feedback in the Tower of Hanoi, helped mitigate Q1 overestimation.

5.2 Analysis of findings

Chapter 2's summary of the previous literature and its perspective on the DKE, and Chapter 3 and 4's continuing and original investigations into the DKE (see Table 9), are considerably varied in their scope, purpose, and results. Analyzing their overall findings therefore allows for determination of the mechanisms involved throughout relative self-assessment. Combining the results obtained in the previous Chapters, suggests the influence (or lack of influence) of five DKE mechanisms; namely egocentrism, task efficacy, chronic self-views, task specific skill, and metacognition.

Table 9: Task specific results and characteristics as per Chapter 3 and 4 findings.

		Sprint	Vertical Jump	Stroop	Tower of Hanoi
Overall estimates	Linear trend	Positive	Positive	Negative	Positive
	Pre-task	Below Average	Above Average	Above Average	Below Average
	Post-task	Above Average	Above Average	Above Average	Below Average
Estimation-performance alignment	Pre-task	0.43	ns	ns	ns
	Post-task	ns	ns	-0.37	0.29
Estimation-performance error	Q1 pre-task	Overestimate	Overestimate	Overestimate	Overestimate
	Q4 pre-task	Underestimate	Underestimate	Underestimate	Underestimate
	Q1 post-task	Overestimate	Overestimate	Overestimate	ns
	Q4 post-task	Underestimate	Underestimate	Underestimate	Underestimate
	Q1 v Q4	Q1 = Q4	Q1 = Q4	Q1 > Q4	Q1 = Q4
	Q1 v Q4	Q1 = Q4	Q1 = Q4	Q1 > Q4	Q1 < Q4
Estimation moderation	Pre-Post	Increase	ns	Increase	Decrease
	Stu-Ath	Decrease	Decrease	Decrease	ns
	Q1 Pre-Post	ns	ns	Increase	Decrease
	Q1 Stu-Ath	ns	ns	ns	ns
	Q4 Pre-Post	ns	ns	Increase	Decrease
	Q4 Stu-Ath	ns	Decrease	ns	ns
Task experience	Estimate	ns	ns	No experience	No experience
	Error	ns	ns	No experience	No experience
Task efficacy	Q1 v Q4 pre	un	un	Q1 = Q4	Q1 = Q4
	Q1 v Q4 post	un	un	Q1 > Q4	Q1 = Q4
	Q1 pre-post	un	un	Increase	Decrease
	Q4 pre-post	un	un	Increase	Decrease
	High predicts	un	un	Overestimation	Overestimation
	Low predicts	un	un	Underestimation	Underestimation
Task characteristics	Feedback	Moderate	Low	Low	High
	Difficulty	Low	Low	Low	High
	Familiarity	High	High	Low	Low

Note: ns = non-significant; Increase/Decrease = indicate direction of change for each specific variable; un = unknown due to exclusion from methodology; High/Moderate/Low = indicates level at which task displayed specific characteristic.

5.2.1 Egocentrism

Egocentrism occurs in comparative judgments of performance capability when individuals place more weight on their own performance perceptions, rather than on those they are referencing against (Battistelli et al., 2009; Kruger, 1999). This can lead individuals to dismiss the performances of others, showing estimation adjustment due to perceived task difficulty (Burson et al., 2006; Kruger, 1999), and a lack of adjustment according to reference

group (Moore & Cain, 2007). If so, egocentrism would be the cause of both better-than-average (Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995) and worse-than-average effects (Kruger, 1999; Moore, 2007), throughout performance estimations.

However, if egocentric weighting provided a strong contribution to estimation (in)accuracy, then individuals would consistently show minimal estimation adjustment when changing the reference group from student to athlete (e.g., Tower of Hanoi). Although participants in Chapters 3 & 4 were, for the most part, unable to show significant quartile specific adjustment (except Vertical Jump; see Table 9), they were able to show an overall adjustment independent of quartile, suggesting participants were able to take others performances into account. Further, although manipulating task difficulty in Chapter 4 led to adjustments in estimations similar to Burson et al. (2006), they alternatively suggest the influence of task efficacy and chronic self-views.

5.2.2 Task efficacy

Efficacy is the belief in one's ability to perform a task (Bandura, 1977), with higher levels of efficacy indicating increased confidence in the ability to successfully perform. This belief has been shown to significantly predict estimation error (Chapter 4) and influence actual performance (Bouffard-Bouchard, Parent, & Larivee, 1991; Coutinho, 2008); and be acquired through previous performance attempts (e.g., chronic self-views; Bandura, 1982; Ehrlinger & Dunning, 2003; Gist, 1987), as well as through immediate task feedback (Chapter 4). As shown in Chapter 4, tasks that appear easy seem to increase an individual's task efficacy, and lead to performance overestimation; whereas more difficult tasks seem to decrease efficacy, leading to performance underestimation (see Table 9). While efficacy was not recorded in Chapter 3 due to it being beyond the scope of Chapter 3's purpose, the raise in performance estimates post-task could have been due to a rise in task efficacy, influenced by both a perceived successful completion (Elzubeir & Rizk, 2000) and increased perceived

task effort (Taras, 2001). As higher efficacy seems to lead to both higher actual performance and overestimation, which are contrasting qualities in DKE performers, additional moderation may occur. This moderation could be due to chronic self-views, as well as metacognition.

5.2.3 Chronic self-views

While efficacy judgments exist in relation to an individual's specific performance or outcome, chronic self-views consist of an assimilation of previously constructed perceptions in order to produce a current estimation of performance capability (Ehrlinger & Dunning, 2003). This assimilation may incorporate both previous experiences in the identical performance, or be based on performances in contextually similar tasks. Without actual performance to influence estimates, chronic self-views are the main source of information used to predict performance prior to task participation. Therefore chronic self-views can be obtained through the use of pre-performance estimates. Alternatively to egocentrism, this would suggest that better-than-average pre-performance estimates (e.g., Vertical Jump, Stroop) are due to more optimistic self-views, whereas worse-than-average pre-performance estimates (e.g., Sprint, Tower of Hanoi) are due to more pessimistic or cautious self-views (Chapter 4).

Chronic self-views have also been shown to impact post-task performance estimates (Ehrlinger & Dunning, 2003); through the influence of performance feedback on psychosocial characteristics such as task efficacy (Chapter 4). Optimistic self-views, along with more positive performance feedback, induce a rise in task efficacy leading to increased post-performance estimates (i.e., Sprint, Stroop). While individuals are more accepting of positive feedback compared to negative feedback (Ditto & Lopez, 1992), pessimistic self-views along with more negative feedback also induce a lowering of task efficacy, leading to decreased post-performance estimates (i.e., Tower of Hanoi). The lack of feedback in the Vertical

Jump, may have neither supported nor refuted original chronic self-views, leaving performance estimations according to timing stable.

Together, this explains why task feedback indicating a low difficulty (e.g., successful completion) leads to increases in task efficacy and higher performance estimates (i.e., overestimation); while task feedback indicating a high difficulty (e.g., task failure) leads to decreases in task efficacy and lower performance estimates (i.e., underestimation; Chapter 4). Alternatively, this explains the average pre and post-task estimations in Chapter 3, as well as the effect of task difficulty on self-assessments previously attributed to individual egocentrism (Burson et al., 2006; Kruger, 1999).

5.2.4 Task specific skill

The direction and degree of estimation error has also largely been attributed to the actual performance capability of individuals (Kruger & Dunning, 1999), with bottom performers (Q1) consistently overestimating actual performance, while top performers (Q4) consistently underestimating it (e.g., Burson et al., 2006; Ehrlinger et al., 2008; Krueger & Mueller, 2002). As research has indicated however, performance estimations are not influenced by actual performance (Ehrlinger & Dunning, 2003). The link between estimated and actual performance may instead be mediated by metacognition, as increases in metacognitive skill are linked to increased actual performance (Coutinho, 2008; Kruger & Dunning, 1999).

While portraying estimation error according to quartile of performance allows for easier understanding of the DKE, characterizing it as a mechanism of the DKE suggests that interventions limited to increasing task specific skill will alleviate estimation error. However as the DKE is a pervasive cognitive bias that affects individuals regardless of actual performance (Chapter 2; Kruger & Dunning, 1999), increasing individual skill will only change the direction of estimation error (i.e., from overestimation to underestimation).

Further, as performance in the DKE is relative to other performers, estimations of performance capability will always be highly influenced by the performance of others.

5.2.5 Metacognition

Metacognition refers to one's knowledge regarding their cognition and cognitive processes (Flavell, 1979), where metacognitive skill is the ability to monitor and control these processes. Originally theorized by Kruger and Dunning (1999) as the main DKE mechanism behind estimation inaccuracy in Q1 performers, metacognition has been consistently linked to accurate self-assessments (Ehrlinger et al., 2008) and actual performance (Coutinho, 2007; Romainville, 1994) in various contexts (e.g., cognitive, physical) as well as performers (e.g., Q1, Q4)

Although multiple strategies have been determined to increase metacognition, such as increased feedback and goal setting (Butler et al., 2008; Zimmerman, 1990), limited studies have attempted to utilize this information to increase individual metacognitive skill (Chapter 2). The use of a metacognitive training packet (Kruger & Dunning, 1999), and increased ongoing task feedback (Chapter 4), serves as the only successful attempts to significantly increase estimation accuracy throughout the DKE literature.

The positive affects of metacognition on estimation accuracy however are quite widespread, influencing not only actual performance, but also task efficacy (Coutinho, 2008), enabling metacognition to mediate the effects of chronic self-views, egocentrism, and performance feedback have on performance estimations. The ability of metacognition to potentially influence every aspect and mechanism of self-assessment bears substantial consideration, especially when attempting to increase individual estimation accuracy regardless of context.

5.3 Statement of Originality

Together, these three studies significantly expand the understanding of cognitive bias specific to the DKE, supporting previous research describing the DKE as a pervasive and thorough bias, which can significantly influence self-assessment (in)accuracy in a variety of performance contexts. The ability of these Chapters to further DKE understanding is in part due to the exploration of the DKE in unique domains, while also using unique methodological designs.

Although previous meta-analyses have been conducted regarding self-assessment error (e.g., Freund & Kasten, 2012; Mabe & West, 1982), Chapter 2 detailed the first systematic review and meta-analysis conducted specifically investigating the DKE and comparative estimation error. As such, Chapter 2's summary of the current state of DKE research leading to a comprehensive analysis of current DKE methodologies, examined contexts, existing mechanisms, and research gaps, are in itself unique. Further, this study was able to quantitatively analyse DKE studies to determine the individual, methodological, and contextual variations in research methodology, which led to significant modification of individual performance estimation accuracy.

Chapter 3 was the first study to specifically investigate the DKE in physical tasks (i.e., Sprint and Vertical Jump). Using uniquely designed methodology, such as manipulation of the timing of self-estimates (i.e., pre v post-task), and the population group used for comparison (i.e., student v athlete), also allowed for the investigation of DKE inconsistencies within the literature. This specific and targeted study design allowed for a unique analysis into the DKE and its mechanisms in novel contexts, while continuing exploration of the overall cognitive processes involved in cognitive bias and estimation inaccuracy.

Although similar to Chapter 3 in study design, Chapter 4's main contribution was due to the use of unique and unfamiliar cognitive tasks for participant performance (i.e., Stroop

and Tower of Hanoi). These tasks allowed for the manipulation of task difficulty and feedback, while documenting associated changes in individual task efficacy. Through a meticulous analysis of findings, this study potentially determined additional mechanisms for the DKE, specifically chronic self-views, task efficacy, and egocentrism.

5.4 Implications

The unique design and findings in the discussed Chapters allow for a detailed determination of the potential implications from (in)accurate self-assessments, of which both Q1 and Q4 estimation error are considerable. Concerning Q1 performers for example, overestimation can result in: increased health risk from smoking related diseases (Lee, 1989); financial loss (Ferraro, 2010); laboratory mistakes (Haun et al., 2000); property and personal loss due to unsafe driving (De Craen et al., 2011; Jonah, 1986); patient death (Whitaker, 2008); decreased student performance (Sanders & Rivers, 1996); and individual injury and death during recreational activities (Palmer, 2002; Petrass et al., 2012). Though less investigated, Q4 underestimation can also lead to significant implications for individual learning (Cury et al., 1997), academic performance (Mattern et al., 2010), and career choice (Ehrlinger et al., 2008).

In spite of the large array of negative consequences to estimation error, very little research has been done to develop interventions designed to mitigate estimation inaccuracy in the DKE. While some studies have attempted to increase self-assessment inaccuracy through increasing post-performance feedback (e.g., Moore & Cain, 2007) and incentivizing accuracy (e.g., Ehrlinger et al., 2008), these have been unsuccessful. Research however, has identified the effectiveness of metacognitive training techniques to mediate estimation error, regardless of performance quartile (Kruger & Dunning, 1999). Additionally, increases in ongoing performance feedback have been shown to indirectly reduce estimation error through its

facilitation of metacognitive monitoring (Butler et al., 2008; Chapter 4; Miller & Geraci, 2011).

While this thesis does not seek to directly decrease estimation error, in order to improve self-assessment accuracy, both this thesis and research in psychology and education literatures collaboratively indicate that increasing metacognitive practices may be the most effective way to improve self-assessment accuracy. To do this, multiple metacognitive building strategies have been theorized, including: planning, goal setting, organization, self-monitoring, environmental structuring, giving self-consequences, rehearsing and memorizing, seeking social assistance, increasing performance feedback, information seeking, record keeping, reviewing, and self-evaluating (Chapter 2; Butler et al., 2008; Schraw, 1998; Zimmerman, 1990). However, while the consequences of inaccuracy are considerable, little research has been done to create a standardized metacognitive building program.

In lieu of this, recommendations still exist for individuals in various professions and circumstances. As metacognition is largely unknown by the public (Hartman, 2001), educators must strive to incorporate metacognitive strategies into lesson plans and lectures. Planning and organizing of course material to reflect course goals is important, along with effective communication of this to students. Course material should be constructed to incorporate easily interpretable performance feedback, so as to encourage students to self-assess, monitor, and evaluate learning throughout the course. Students themselves are advised to keep organized study notes, plan their studying, constantly evaluate their current knowledge, and seek assistance and feedback from other students and educators.

To safeguard company profits and individual safety, self-reflection of individual work safety and efficiency should be encouraged through constant evaluation of individual working/operating practices and environmental safety hazards, accurate and up-to-date record keeping, and self-monitoring of individual fatigue and mental state. Outside the workplace,

individuals should also constantly self-monitor their own personal behaviour, diminish autonomy of daily activities, evaluate behavioural decisions and accompanied actions, and objectively assess performance in relation to others. Regardless of context, these strategies then lead to increased thinking, problem-solving, learning, and knowledge retention (Bransford, Sherwood, Vye, & Rieser, 1986; Hartman, 2001), improving self-assessment accuracy, and diminishing the negative consequences of inaccuracy.

5.5 Limitations

Although the previous chapters provide a unique and significant contribution to the current DKE literature, limitations exist regarding their methodology and scope. While the meta-analysis in Chapter 2 was both timely and important, the exclusion of unpublished studies in the selection criteria may have been limiting. While the various analyses performed indicated no publication bias, and the lack of unpublished material was justified (Baumeister et al., 2009; Ferguson, 2010), using both published and unpublished material (Stroup et al., 2000) may have led to the inclusion of potentially informative studies. As with other systematic reviews and meta-analyses, any information that was initially missed in Chapter 2 would have also affected Chapters 3 and 4, due to Chapter 2 being used as the base of knowledge from which to expand from.

While the main focus of this thesis was the investigation of the DKE, any in-depth analysis of a cognitive bias requires a similar investigation into its mechanism. Throughout Chapters 2 - 4, metacognition is professed as the main mechanisms behind DKE specific cognitive bias and estimation error. While the literature does provide considerable support for this claim, Chapter 2 was unable to find numerous studies that investigated the relationship between metacognition and self-assessment accuracy. Considerable support was only found for the DKE trend itself, leading to conclusion that metacognition is the main mechanism of

the DKE. Due to this support, Chapters 3 and 4 did not directly investigate metacognition, instead restricting the investigation into additional mechanisms such as psycho-social and methodological characteristics. Including metacognition would have allowed for a much more robust determination of the cognitive process involved in self-assessment and their relationship to each other. It would have also provided more information for a general metacognitive building intervention.

This thesis also restricted itself to the analysis of bottom (Q1) and top (Q4) performers and their difficulties accurately aligning estimations with actual performance. As these groups are the most inaccurate, it is advantageous to investigate the cognitive mechanisms leading to their inaccuracy. This allows for future interventions to remedy the most inaccurate.

5.6 Future Directions

Directly diminishing individual estimation error was beyond the scope of this thesis. It is however, able to provide a substantial degree of information regarding the specific mechanisms various intervention strategies may utilize when seeking to improve self-assessment accuracy. Additionally, it has determined which previous interventions have been unsuccessful. The pertinent next step would therefore be to directly test these mechanisms on individual estimation error, seeking to determine the most efficient way to improve self-assessment accuracy.

A considerable degree of further research is necessary to fully determine the cognitive processes ongoing throughout self-assessment, foremost of which are the limitations inherent in the previously discussed Chapters. The investigation into the direct relationship between metacognitive skill, chronic self-views, task efficacy, and other moderators (e.g., psycho-social characteristics) seems important and necessary. In particular, psycho-social

characteristics such as narcissism and modesty are important due to the considerable degree of moderation they seem to exhibit on estimation errors. The inclusion of specific questionnaires that enable investigation of these characteristics into future DKE research would enable determination of their effect. While the investigation into Q2 and Q3 performers was not included in the scope of this thesis, further analysis of the factors influencing their degree of misalignment, and the implications specific to their estimation error would assist in gaining an overall picture of self-assessment capability. Analysing this data through the comparison of mean differences, similar to previous DKE research, would enable a thorough understanding of their estimation error.

To further build upon the findings of this thesis, additional studies investigating the processes behind individual use of performance cues and feedback to self-assess performance are worthwhile. More long-term studies designed to identify the resultant behavioural, learning, functionality, and health consequences of consistent estimated performance (in)accuracies in specific tasks and across various contexts are also necessary. Addressing these gaps will allow for the designing and testing of improved interventions aiming to increase overall estimated performance accuracy, while simultaneously reducing the negative consequences associated with it.

Chapter 6 - Conclusion

This thesis was designed to uniquely examine cognitive bias through the use of a meta-analysis and two investigative studies, allowing for both a descriptive and analytical discussion specifically regarding the DKE. Throughout the literature, a small-moderate capability to align estimations with actual performance was observed ($r = .32$), regardless of individual age, gender, and status; while bottom performers (Q1) were found to underestimate (by 37 percentile points), and top performers (Q4) were found to overestimate actual performance (by -20 percentile points).

Investigating self-assessment accuracy in familiar physical tasks displayed higher estimation-performance alignment compared to unfamiliar cognitive tasks, with consistent small-moderate capability to accurately align estimations with performance (i.e., mean $r = .17$). Q1 performers were shown to overestimate their relative performance (i.e., mean = 43 percentile points), while Q4 performers underestimated their performance (i.e., mean = -29 percentile points); regardless of experience. Manipulation of methodological factors such as the timing of estimations and reference group, and moderator variables such as task context, difficulty and feedback did result in significant changes in estimation accuracy. Psycho-social characteristics such as task efficacy were found to significantly predict estimation error, with low efficacy predicting underestimation, and high efficacy predicting overestimation.

Overall, estimation accuracy seems highly influenced by metacognitive skill, task feedback, and chronic self-views. Increased individual metacognitive skill and increased ongoing performance feedback seems the most effective strategies to mitigate the DKE. The findings of this thesis provide substantial information regarding the cognitive processes during self-assessment, and will contribute significantly to the production of a standardized intervention designed to mitigate estimation error. Further research into the effectiveness of

general metacognitive building techniques, as well as continued research into the cognitive process during self-assessment is needed.

References

(Note: * = included in Chapter 2 meta-analysis, # = Reference number for Chapter 2 systematic review)

- Abramson, L. Y., Seligman, M. E., & Teasdale, J. D. (1978). Learned helplessness in humans: critique and reformulation. *Journal of Abnormal Psychology, 87*(1), 49.
- Albanese, M., Dottl, S., Mejicano, G., Zakowski, L., Seibert, C., Van Eyck, S., & Prucha, C. (2006). Distorted perceptions of competence and incompetence are more than regression effects. *Advances in Health Sciences Education: Theory and Practice, 11*(3), 267-278. doi: 10.1007/s10459-005-2400-7 ¹⁵
- Alicke, M. D., Klotz, M., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology, 68*(5), 804-825. doi: <http://dx.doi.org/10.1037/0022-3514.68.5.804>
- *Ames, D. R., & Kammrath, L. K. (2004). Mind-reading and metacognition: Narcissism, not actual competence, predicts self-estimated ability. *Journal of Nonverbal Behavior, 28*(3), 187-209. doi: 10.1023/B:JONB.0000039649.20015.0e ¹⁰
- Arkes, H. R., Faust, D., Guilmette, T. J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology, 73*(2), 305.
- Arsenian, S. (1942). Own estimate and objective measurement. *Journal of Educational Psychology, 33*(4), 291-302. doi: 10.1037/H0063257 ¹
- Austin, Z., & Gregory, P. A. (2007). Evaluating the accuracy of pharmacy students' self-assessment skills. *American Journal of Pharmaceutical Education, 71*(5), 89.
- *Bailey, K. G., & Lazar, J. (1976). Accuracy of self-ratings of intelligence as a function of sex and level of ability in college-students. *Journal of Genetic Psychology, 129*(2), 279-290. ³

- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37(2), 122.
- Battistelli, P., Cadamuro, A., Farneti, A., & Versari, A. (2009). Do university students know how they perform? *International Electronic Journal of Elementary Education*, 2(1), 180-198. ²¹
- Baumeister, R. F., DeWall, C. N., & Vohs, K. D. (2009). Social Rejection, Control, Numbness, and Emotion How Not To Be Fooled by Gerber and Wheeler (2009). *Perspectives on Psychological Science*, 4(5), 489-493.
- Baxley, S. G., Brown, S. T., Pokorny, M. E., & Swanson, M. S. (1997). Perceived competence and actual level of knowledge of diabetes mellitus among nurses. *Journal of Nursing Staff Development*, 13(2), 93-98.
- Biostat. (2005). *Comprehensive Meta Analysis (Version 2.0)*. Englewood NJ.
- Bobbert, M. F., Gerritsen, K. G., Litjens, M. C., & Van Soest, A. J. (1996). Why is countermovement jump height greater than squat jump height? *Medicine and Science in Sports and Exercise*, 28, 1402-1412.
- Bond Jr, C. F., Wiitala, W. L., & Richard, F. D. (2003). Meta-analysis of raw mean differences. *Psychological Methods*, 8(4), 406.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009) *Introduction to Meta-Analysis*: John Wiley & Sons, Ltd.
- Bouffard-Bouchard, T., Parent, S., & Larivee, S. (1991). Influence of self-efficacy on self-regulation and performance among junior and senior high-school age students. *International Journal of Behavioral Development*, 14(2), 153-164.
- Bransford, J., Sherwood, R., Vye, N., & Rieser, J. (1986). Teaching thinking and problem solving: Research foundations. *American Psychologist*, 41(10), 1078.

- *Brim, O. G., Jr. (1954). College grades and self-estimates of intelligence. *Journal of Educational Psychology*, 45(8), 477-484. doi: 10.1037/h0057492 ²
- *Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology*, 90(1), 60-77. doi: 10.1037/0022-3514.90.1.60 ¹³
- Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2008). Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 918.
- Butler, R. (1990). The Effects of Mastery and Competitive Conditions on Self-Assessment at Different Ages. *Child Development*, 61(1), 201-210.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Routledge.
- Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, 4(03), 317-331.
- Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis*: Russell Sage Foundation.
- Coutinho, S. (2008). Self-efficacy, metacognition, and performance. *North American Journal of Psychology*, 10(1), 165.
- Coutinho, S. A. (2007). The relationship between goals, metacognition, and academic success. *Educate*, 7(1), 39-47.
- Cox, J. R., & Griggs, R. A. (1982). The effects of experience on performance in Wason's selection task. *Memory & Cognition*, 10(5), 496-502.

- Critcher, C. R., & Dunning, D. (2009). How chronic self-views influence (and mislead) self-assessments of task performance: self-views shape bottom-up experiences with the task. *Journal of Personality and Social Psychology, 97*(6), 931.
- Cross, K. P. (1977). Not can, but will college teaching be improved? *New Directions for Higher Education, 1977*(17), 1-15. doi: 10.1002/he.36919771703
- CSEP. (2002). PAR-Q & You. *Canadian Society for Exercise Physiology*.
- Cury, F., Biddle, S., Sarrazin, P., & Famose, J. P. (1997). Achievement goals and perceived ability predict investment in learning a sport task. *British Journal of Educational Psychology, 67*(3), 293-309.
- *De Craen, S., Twisk, D. A., Hagenzieker, M. P., Elffers, H., & Brookhuis, K. A. (2011). Do young novice drivers overestimate their driving skills more than experienced drivers? Different methods lead to different conclusions. *Accident; Analysis and Prevention, 43*(5), 1660-1665. ²⁴
- DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*(3), 177-188. doi: dx.doi.org/10.1016/0197-2456(86)90046-2
- Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology, 63*(4), 568.
- Duda, J. L., & Nicholls, J. G. (1992). Dimensions of Achievement-Motivation in Schoolwork and Sport. *Journal of Educational Psychology, 84*(3), 290-299. doi: 10.1037/0022-0663.84.3.290
- Dunning, D. (2011). The dunning-kruger effect: On being ignorant of one's own ignorance *Advances in Experimental Social Psychology* (Vol. 44, pp. 247-296). San Diego, CA: Elsevier.

- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455-463.
- *Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology*, *84*(1), 5-17. ⁸
- *Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes*, *105*(1), 98-121. doi: 10.1016/j.obhdp.2007.05.002 ²⁰
- Elzubeir, M., & Rizk, D. (2000). Assessing confidence and competence of senior medical students in an obstetrics and gynaecology clerkship using an OSCE. *Education for Health* *14*(3), 373-382.
- Epp, A. M., Dobson, K. S., Dozois, D. J., & Frewen, P. A. (2012). A systematic meta-analysis of the Stroop task in depression. *Clinical Psychology Review*, *32*(4), 316-328. doi: 10.1016/j.cpr.2012.02.005
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The Role of Deliberate Practice in the Acquisition of Expert Performance. *Psychological Review*, *100*(3), 363-406. doi: Doi 10.1037/0033-295x.100.3.363
- Eva, K. W., Cunnington, J. P., Reiter, H. I., Keane, D. R., & Norman, G. R. (2004). How can I know what I don't know? Poor self assessment in a well-defined domain. *Advances in Health Science Education: Theory and Practice*, *9*(3), 211-224. doi: 10.1023/B:AHSE.0000038209.65714.d4
- Felson, R. B. (1981). Ambiguity and bias in the self-concept. *Social Psychology Quarterly*, *44*(1), 64-69. doi: 10.2307/3033866

- Feltz, D. L. (1988). Self-confidence and sports performance. *Exercise and Sport Sciences Reviews, 16*, 423-457.
- Ferguson, C. J. (2010). A meta-analysis of normal and disordered personality across the life span. *Journal of Personality and Social Psychology, 98*(4), 659-667. doi: 10.1037/a0018770
- *Ferraro, P. J. (2010). Know thyself: Competence and self-awareness. *Atlantic Economic Journal, 38*(2), 183-196. ²²
- Fitzgerald, J. T., White, C. B., & Gruppen, L. D. (2003). A longitudinal study of self-assessment accuracy. *Medical Education, 37*(7), 645-649.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906-911. doi: dx.doi.org/10.1037/0003-066X.34.10.906
- Freelap Australia. (2012). Fly Sprint Kit. Aberfoyle Park, SA: Reverb.
- Freund, P. A., & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin, 138*(2), 296-321. doi: dx.doi.org/10.1037/a0026556
- *Furnham, A. (2005). Self-Estimated Intelligence, Psychometric Intelligence and Personality. *Psychologia - An International Journal of Psychology in the Orient, 48*(3), 182-192. doi: 10.2117/psysoc.2005.182 ¹¹
- *Furnham, A., & Fong, G. (2000). Self-estimated and psychometrically measured intelligence: A cross-cultural and sex differences study of British and Singaporean students. *North American Journal of Psychology, 2*(2), 191-200. ⁵
- Furnham, A., Kidwai, A., & Thomas, C. (2001). Personality, psychometric intelligence, and self-estimated intelligence. *Journal of Social Behavior and Personality, 16*(1), 97-114.

- *Furnham, A., Zhang, J., & Chamorro-Premuzic, T. (2006). The relationship between psychometric and self-estimated intelligence, creativity, personality and academic achievement. *Imagination, Cognition and Personality, 25*(2), 119-145. doi: 10.2190/530v-3m9u-7uq8-fmbg ¹⁴
- Gati, I., Fishman-Nadav, Y., & Shiloh, S. (2006). The relations between preferences for using abilities, self-estimated abilities, and measured abilities among career counseling clients. *Journal of Vocational Behavior, 68*(1), 24-38. doi: dx.doi.org/10.1016/j.jvb.2005.04.002
- Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17*(3), 295-314.
- Gist, M. E. (1987). Self-efficacy: Implications for organizational behavior and human resource management. *Academy of Management Review, 12*(3), 472-485.
- Goel, V., & Grafman, J. (1995). Are the frontal lobes implicated in "planning" functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia, 33*(5), 623-642.
- Goldberg, T. E., Saint-Cyr, J. A., & Weinberger, D. (1990). Assessment of procedural learning and problem solving in schizophrenic patients by Tower of Hanoi type tasks. *Journal of Neuropsychiatry and Clinical Neurosciences, 2*(2), 165-173.
- Greifeneder, R., Bless, H., & Pham, M. T. (2010). When do people rely on affective and cognitive feelings in judgment? A review. *Personality and Social Psychology Review, 15*(2), 107-141.
- *Gross, M., & Latham, D. (2007). Attaining information literacy: An investigation of the relationship between skill level, self-estimates of skill, and library anxiety. *Library & Information Science Research, 29*(3), 332-353. doi: 10.1016/j.lisr.2007.04.012 ¹⁷
- *Gross, M., & Latham, D. (2012). What's skill got to do with it?: Information literacy skills and self-views of ability among first-year college students. *Journal of the American*

Society for Information Science and Technology, 63(3), 574-583. doi:

10.1002/asi.21681 ²⁵

Gutin, B., Humphries, M. C., Lemmon, C. R., Waller, J. L., Baranowski, T., Rogers, L. Q., & Saunders, R. (2006). Evaluation of Internal Medicine Residents as Exercise Role Models and Associations With Self-Reported Counseling Behavior, Confidence, and Perceived Success. *Teaching and Learning in Medicine*, 18(3), 215-221. doi:

10.1207/s15328015tlm1803_5

Hartman, H. J. (2001). Developing students' metacognitive knowledge and skills

Metacognition in learning and instruction (pp. 33-68): Springer.

Haun, D. E., Zeringue, A., Leach, A., & Foley, A. (2000). Assessing the competence of specimen-processing personnel. *Laboratory Medicine*, 31(11), 633-637. ⁶

Heider, F. (2013). *The psychology of interpersonal relations*. New York: Psychology Press.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis.

Statistics in Medicine, 21(11), 1539-1558. doi: 10.1002/sim.1186

IBM Corp. (2012). IBM SPSS Statistics for Windows. Armonk, NY: IBM Corp.

Jacsó, P. (2005). Google Scholar: the pros and the cons. *Online information review*, 29(2), 208-214.

Jacsó, P. (2008). The pros and cons of computing the h-index using Google Scholar. *Online information review*, 32(3), 437-452.

Jahoda, M. (1958). *Current concepts of positive mental health* (Vol. 1). New York, NY: Basic Books.

*Jaramillo, F., Carrillat, F. A., & Locander, W. B. (2003). Starting to solve the method puzzle in salesperson self-report evaluations. *Journal of Personal Selling & Sales Management*, 23(4), 369-377. ⁹

- Jonah, B. A. (1986). Accident risk and risk-taking behaviour among young drivers. *Accident Analysis & Prevention, 18*(4), 255-271.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*(3), 430-454. doi: [http://dx.doi.org/10.1016/0010-0285\(72\)90016-3](http://dx.doi.org/10.1016/0010-0285(72)90016-3)
- Kaufman, J. C., Evans, M. L., & Baer, J. (2010). The American Idol Effect: Are students good judges of their creativity across domains? *Empirical Studies of the Arts, 28*(1), 3-17.
- Keith, N., & Frese, M. (2005). Self-regulation in error management training: emotion control and metacognition as mediators of performance effects. *Journal of Applied Psychology, 90*(4), 677.
- Kistler, B. M., Walsh, M. S., Horn, T. S., & Cox, R. H. (2010). The acute effects of static stretching on the sprint performance of collegiate men in the 60-and 100-m dash after a dynamic warm-up. *The Journal of Strength & Conditioning Research, 24*(9), 2280-2284.
- Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology, 17*(2), 248-294.
- *Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality and Social Psychology, 82*(2), 180-188. doi: 10.1037//0022-3514.82.2.180 ⁷

- Kruger, J. (1999). Lake Wobegon be gone! The "Below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology*, 77(2), 221-232. doi: Doi 10.1037//0022-3514.77.2.221
- *Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121-1134. doi: 10.1037//0022-3514.77.6.1121 ⁴
- Kruger, J., & Dunning, D. (2002). Unskilled and unaware - But why? A reply to Krueger and Mueller (2002). *Journal of Personality and Social Psychology*, 82(2), 189-192.
- Lafayette Instrument. (2004). Vertisonic (Version 01100). Lafayette, IN: Lafayette Instrument Company.
- Lee, C. (1989). Perceptions of immunity to disease in adult smokers. *Journal of Behavioral Medicine*, 12(3), 267-277.
- Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, 67(3), 280-296. doi: 10.1037/0021-9010.67.3.280
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin*, 109(2), 163-203.
- Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, 1(2), 161-175.
- Marcora, S. (2009). Perception of effort during exercise is independent of afferent feedback from skeletal muscles, heart, and lungs. *Journal of Applied Physiology*, 106(6), 2060-2062.

- *Mattern, K. D., Burrus, J., & Shaw, E. (2010). When both the skilled and unskilled are unaware: Consequences for academic performance. *Self and Identity*, 9(2), 129-141. doi: 10.1080/15298860802618963 ²³
- Merkle, C., & Weber, M. (2011). True overconfidence: The inability of rational information processing to account for apparent overconfidence. *Organizational Behavior and Human Decision Processes*, 116(2), 262-271.
- *Mikkelsen, L., Kaprio, J., Kautiainen, H., Kujala, U. M., & Nupponen, H. (2005). Associations between self-estimated and measured physical fitness among 40-year-old men and women. *Scandinavian Journal of Medicine & Science in Sports*, 15(5), 329-335. doi: 10.1111/j.1600-0838.2004.00429.x ¹²
- Miller, T. M., & Geraci, L. (2011). Training metacognition in the classroom: The influence of incentives and feedback on exam predictions. *Metacognition and Learning*, 6(3), 303-314.
- Mishra, S. (2014). Decision-Making Under Risk: Integrating Perspectives From Biology, Economics, and Psychology. *Personality and Social Psychology Review*, 18(3), 280-307. doi: 10.1177/1088868314530517
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, 151(4), 264-269.
- Moore, D. A. (2007). Not so above average after all: When people believe they are worse than average and its implications for theories of bias in social comparison. *Organizational Behavior and Human Decision Processes*, 102(1), 42-58.
- Moore, D. A., & Cain, D. M. (2007). Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition. *Organizational Behavior and Human Decision Processes*, 103(2), 197-213. ¹⁸

- *Moore, D. A., & Small, D. A. (2007). Error and bias in comparative judgment: On being both better and worse than we think we are. *Journal of Personality and Social Psychology*, 92(6), 972-989. doi: 10.1037/0022-3514.92.6.972 ¹⁹
- Mullen, B., Atkins, J. L., Champion, D. S., Edwards, C., Hardy, D., Story, J. E., & Vanderklok, M. (1985). The false consensus effect: A meta-analysis of 115 hypothesis tests. *Journal of Experimental Social Psychology*, 21(3), 262-283.
- Nadeau, R., Cloutier, E., & Guay, J.-H. (1993). New evidence about the existence of a bandwagon effect in the opinion formation process. *International Political Science Review*, 14(2), 203-213.
- Palmer, C. (2002). 'Shit happens': The selling of risk in extreme sport. *The Australian Journal of Anthropology*, 13(3), 323-336.
- Pazicni, S., & Bauer, C. F. (2014). Characterizing illusions of competence in introductory chemistry students. *Chemistry Education Research and Practice*, 15(1), 24-34.
- Petrass, L. A., Blitvich, J. D., McElroy, G. K., Harvey, J., & Moran, K. (2012). Can you swim? Self-report and actual swimming competence among young adults in ballarat, australia. *International Journal of Aquatic Research & Education*, 6(2), 136-148.
- Phillips, L. H., Bull, R., Adams, E., & Fraser, L. (2002). Positive mood and executive function: evidence from stroop and fluency tasks. *Emotion*, 2(1), 12.
- Pothos, E. M., & Tapper, K. (2010). Inducing a Stroop effect. *Applied Cognitive Psychology*, 24(7), 1021-1033.
- Romainville, M. (1994). Awareness of cognitive strategies: The relationship between university students' metacognition and their performance. *Studies in Higher Education*, 19(3), 359-366.

- Rösch, D., Hodgson, R., Peterson, L., Graf-Baumann, T., Junge, A., Chomiak, J., & Dvorak, J. (2000). Assessment and evaluation of football performance. *The American Journal of Sports Medicine*, 28(suppl 5), S-29-S-39.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638-641. doi: 10.1037//0033-2909.86.3.638
- Ross, L., Greene, D., & House, P. (1977). The “false consensus effect”: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 13(3), 279-301.
- *Ryvkin, D., Krajč, M., & Ortmann, A. (2012). Are the unskilled doomed to remain unaware? *Journal of Economic Psychology*, 33(5), 1012-1031. doi: 10.1016/j.joep.2012.06.003 ²⁶
- Sanders, W. L., & Rivers, J. C. (1996). Cumulative and residual effects of teachers on future student academic achievement: Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.
- *Schlösser, T., Dunning, D., Johnson, K. L., & Kruger, J. (2013). How unaware are the unskilled? Empirical tests of the “signal extraction” counterexplanation for the Dunning–Kruger effect in self-evaluation of performance. *Journal of Economic Psychology*, 39(0), 85-100. doi: 10.1016/j.joep.2013.07.004 ²⁷
- Schmidt, A. M., & Ford, J. K. (2003). Learning within a learner control training environment: The interactive effects of goal orientation and metacognitive instruction on learning outcomes. *Personnel Psychology*, 56(2), 405-429.
- Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional science*, 26(1-2), 113-125.

- *Sheldon, O. J., Dunning, D., & Ames, D. R. (2014). Emotionally unskilled, unaware, and uninterested in learning more: Reactions to feedback about deficits in emotional intelligence. *Journal of Applied Psychology, 99*(1), 125. ³⁰
- Steel, K. A., & Eisenhuth, J. (2012). The team-mate identification (TM-ID) test: A portable apparatus for collecting decision latencies for players in team invasion sports. *International Journal of Sports Science and Engineering, 6*(3), 159-164.
- Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., . . . Thacker, S. B. (2000). Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA, 283*(15), 2008-2012.
- Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A., & Song, F. (2000). *Methods for meta-analysis in medical research*. Chichester, UK: J. Wiley.
- Taras, M. (2001). The use of tutor feedback and student self-assessment in summative assessment tasks: Towards transparency for students and for tutors. *Assessment & Evaluation in Higher Education, 26*(6), 605-614.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological Bulletin, 103*(2), 193-210.
- Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher, 31*(3), 25-32.
- Vanyperen, N. W. (1992). Self-enhancement among major league soccer players: The role of importance and ambiguity on social comparison behavior. *Journal of Applied Social Psychology, 22*(15), 1186-1198. doi: 10.1111/j.1559-1816.1992.tb02359.x
- Violato, C., & Lockyer, J. (2006). Self and peer assessment of pediatricians, psychiatrists and medicine specialists: implications for self-directed learning. *Advances in Health Sciences Education: Theory and Practice, 11*(3), 235-244. doi: 10.1007/s10459-005-5639-0 ¹⁶

- *von Stumm, S. (2013). Intelligence, gender, and assessment method affect the accuracy of self-estimated intelligence. *British Journal of Psychology*, *105*(2), 243-253. doi: 10.1111/bjop.12031 ²⁸
- Waylen, A. E., Horswill, M. S., Alexander, J. L., & McKenna, F. P. (2004). Do expert drivers have a reduced illusion of superiority? *Transportation Research Part F: Traffic Psychology and Behaviour*, *7*(4), 323-331.
- Welsh, M. C. (1991). Rule-guided behavior and self-monitoring on the Tower of Hanoi disk-transfer task. *Cognitive Development*, *6*(1), 59-76.
- Welsh, M. C., Satterlee-Cartmell, T., & Stine, M. (1999). Towers of Hanoi and London: contribution of working memory and inhibition to performance. *Brain and Cognition*, *41*(2), 231-242. doi: 10.1006/brcg.1999.1123
- Whitaker, M. (2008). CMO's annual report 2007. *Bulletin of The Royal College of Surgeons of England*, *90*(8), 256-256.
- *Williams, E. F., Dunning, D., & Kruger, J. (2013). The hobgoblin of consistency: algorithmic judgment strategies underlie inflated self-assessments of performance. *Journal of Personality and Social Psychology*, *104*(6), 976-994. doi: 10.1037/a0032416 ²⁹
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, *25*(1), 3-17.

Appendices

Appendix A – PRISMA Checklist

Section/topic	#	Checklist item	Reported on page
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	7
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	7
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	10
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	11
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	12
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	12
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	12
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	12
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	12
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	13
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	13
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	n/a
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	15
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	16
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	16

Appendix A –Continued

RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	18
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	19
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	n/a
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	19
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	25
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	16
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	29
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	31
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	38
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	91
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	n/a

Appendix B – Example list of excluded studies

Title rejection:

Dungen, H. D., Heidelk, J., Kruger, M., Krackhardt, F., & Haverkamp, W. (2007). Beta receptor blocker therapy of heart failure with particular consideration of elderly patients. *Medizinische Welt*, 58(1-2), 5-10.

Halpern, J. H., Sherwood, A. R., Hudson, J. I., Gruber, S., Kozin, D., & G. (2011). Residual neurocognitive features of long-term ecstasy users with minimal exposure to other drugs. *Addiction*, 106(4), 777-786. doi: 10.1111/j.1360-0443.2010.03252.x

'Healthy Penis' campaign targets syphilis risk. (2007). *Contraceptive Technology Update*, 28(4), 44-46.

Abstract rejection:

Berardi-Coletta, B., Buyer, L. S., Dominowski, R. L., & Rellinger, E. R. (1995).

Metacognition and Problem Solving: A Process-Oriented Approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(1), 205-223.

Bipp, T., & Kleingeld, A. (2012). Self-estimates of intelligence: interaction effects of the comparison to a specific reference group and neuroticism. *Psychological Reports*, 110(2), 403-415.

Morley-Hauchecorne, C., & Lepatourel, J. A. (2000). Self-perceived competence of clinical dietitians to participate in research: a needs assessment. *Canadian Journal of Dietetic Practice & Research*, 61(1), 6-12.

Appendix B - *Continued*

Full-text rejection:

Furnham, A. (2004). Are lay people lumpers or splitters? The factor structure of, and sex differences related to, self-rated and other-rated abilities. *Learning and Individual Differences*, 14(3), 153-168. doi: DOI 10.1016/j.lindif.2004.02.001

Kruger, A. (2010). Sport psychological skills that discriminate between successful and less successful female university field hockey players. *African Journal for Physical, Health Education, Recreation & Dance*, 16(2), 240-250.

Kruger, J., & Dunning, D. (2002). Unskilled and unaware - But why? A reply to Krueger and Mueller (2002). *Journal of Personality and Social Psychology*, 82(2), 189-192.

Appendix C – Human Research Ethics Committee Approval



THE UNIVERSITY OF
SYDNEY

Research Integrity
Human Research Ethics Committee

Thursday, 9 May 2013

Dr Stephen Cobley
Exercise Health and Performance; Faculty of Health Sciences
Email: stephen.cobley@sydney.edu.au

Dear Stephen

I am pleased to inform you that the University of Sydney Human Research Ethics Committee (HREC) has approved your project entitled "**The Dunning-Kruger Effect as a Constraint on Learning and Expertise: Insights from Motor & Cognitive Performance Contexts.**".

Details of the approval are as follows:

Project No.: 2013/345
Approval Date: 09/05/2013
First Annual Report Due: 09/05/2014
Authorised Personnel: Cobley Stephen; Adams Roger; Hubka Tate;
Documents Approved:

Date Uploaded	Type	Document Name
08/04/2013	Other Type	Participant Debrief
08/04/2013	Questionnaires/Surveys	Pre/Post Task Questionnaire
03/05/2013	Advertisements/Flyer	Advertisement - Version Two
03/05/2013	Participant Info Statement	Participant Information Statement - Version Two
03/05/2013	Questionnaires/Surveys	Pre-Test Questionnaire - Version Two
03/05/2013	Amendment/Modification form	Response to existing application - Cover Letter

HREC approval is valid for four (4) years from the approval date stated in this letter and is granted pending the following conditions being met:

Condition/s of Approval

- Continuing compliance with the National Statement on Ethical Conduct in Research Involving Humans.
- Provision of an annual report on this research to the Human Research Ethics Committee from the approval date and at the completion of the study. Failure to submit reports will result in withdrawal of ethics approval for the project.
- All serious and unexpected adverse events should be reported to the HREC within 72 hours.

Research Integrity
Research Portfolio
Level 2, Margaret Telfer
The University of Sydney
NSW 2006 Australia

T +61 2 8627 8111
F +61 2 8627 8177
E ro.humanethics@sydney.edu.au
sydney.edu.au

ABN 15 211 513 464
CRICOS 00026A

Appendix C – Continued



- All unforeseen events that might affect continued ethical acceptability of the project should be reported to the HREC as soon as possible.
- Any changes to the project including changes to research personnel must be approved by the HREC before the research project can proceed.

Chief Investigator / Supervisor's responsibilities:

1. You must retain copies of all signed Consent Forms (if applicable) and provide these to the HREC on request.
2. It is your responsibility to provide a copy of this letter to any internal/external granting agencies if requested.

Please do not hesitate to contact Research Integrity (Human Ethics) should you require further information or clarification.

Yours sincerely

Dr Stephen Assinder
Chair
Human Research Ethics Committee

This HREC is constituted and operates in accordance with the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Human Research (2007), NHMRC and Universities Australia Australian Code for the Responsible Conduct of Research (2007) and the CPMP/ICH Note for Guidance on Good Clinical Practice.

Appendix D – Study Advertisement

Psychological Processes in Learning & Performing



We would like to examine some of these psychological processes with you!

To assess some of the various psychological mechanisms, we would like to invite you to have a go at completing some simple, fun, unique, as well as familiar movement and puzzle based tasks.

We are seeking volunteers from Health Sciences students to participate in the study. Participants will have the chance to have a go at several fun and unique tasks. The tasks will take approximately one and a half hours to complete. Physical activity will be involved.

For more information, contact the research team:

Tate Hubka: (p) 9036 7366 (e) tate.hubka@sydney.edu.au

or

Dr. Steve Coble (p) 9351 9033 (e) stephen.coble@sydney.edu.au

Appendix E – Participant Information Statement/Consent Form

Psychological Processes in Learning & Performing

PARTICIPANT INFORMATION STATEMENT

(1) What is the study about?

This study is looking at the different psychological processes that affect learning and performance in a sport context.

(2) Who is carrying out the study?

The study is being conducted by Dr. Steve Cobley, Mr. Tate Hubka, & Dr. Roger Adams. Tate Hubka is conducting this study as part of a Masters Degree.

(3) What does the study involve?

This study involves the completion of a series of questionnaires, and then the completion of two maximum physical exertion tasks, a cognitive task, and a reaction time task. We cannot specify exactly to you what these consist of, but many are brief tasks (e.g., requiring under 10 sec of physical exertion, or approximately 10 minutes for a non physical task), that include performing all kinds of skills or problem-solving.

(4) How much time will the study take?

The whole procedure should take no longer than an hour and a half of your time in order to complete all of the tasks.

(5) Is there any risk involved?

There is no more risk involved than in your activities of daily living.

(6) Can I withdraw from the study?

You can withdraw from this study at any time with no repercussions, prejudice or fear of penalty.

(7) Will anyone else know the results?

Personal information will be known only to the researchers directly involved. Efforts will be made to ensure anonymity and confidentiality when dealing with your results in this study. Results may be published in scientific literature, though no information related to you will be used in a way that is identifiable.

(8) Will the study benefit me?

This study will have no direct benefits to you. Although you may find that the results of the study useful in understanding how your personal psychology and the psychology of others influences learning and performance.

Appendix E – Continued

(9) Can I tell other people about the study?

During data collection of the study, we will ask that you refrain from discussing or informing others specifically about the nature and content of the study, as this may affect the perceptions of future participants. We are happy however for you to share your general experience of being involved in this study (e.g., I enjoyed being involved). After data collection is completed, we are more than happy for you to disclose and share the specifics of the study content, and your experience.

(10) What if I require further information about the study or my involvement in it?

When you have read this information, the researcher will discuss it with you further and answer any questions you may have. If you would like to know more at any stage, please feel free to contact Dr. Steve Copley at 9351 9033; email: stephen.copley@sydney.edu.au or Tate Hubka at 9036 7366; email: tate.hubka@sydney.edu.au

(11) What if I have a complaint or any concerns?

Any person with concerns or complaints about the conduct of a research study can contact The Manager, Human Ethics Administration, University of Sydney on +61 2 8627 8176 (Telephone); +61 2 8627 8177 (Facsimile) or ro.humanethics@sydney.edu.au (Email).

This information sheet is for you to keep.

Appendix E – Continued

PARTICIPANT CONSENT FORM

I,[PRINT NAME], give consent to my participation in the research project, entitled:

Psychological Processes in Learning and Performance

In giving my consent I acknowledge that:

1. The procedures required for the project and the time involved have been explained to me, including any inconvenience, risk, discomfort or side effect, and their implications, and any questions I have about the project have been answered to my satisfaction.
2. I have read the *Participant Information Statement* and have been given the opportunity to discuss the information and my involvement in the project with the researcher/s.
3. I understand that being in this study is completely voluntary – I am not under any obligation to consent.
4. I understand that my involvement is strictly confidential. I understand that any research data gathered from the results of the study may be published however no information about me will be used in any way that is identifiable.
5. I understand that I can withdraw from the study at any time, without affecting my status as a student, nor will it affect any relationship with individuals or the organisations associated with this study.

Initials

Appendix E – Continued

I consent to:

- Completing all questionnaires.
YES NO
- Completing the physical, and cognitive tests.
YES NO

Feedback Option

If you would like to receive feedback on this research and your results, please provide your details (i.e., mailing address, email address); and check the “yes” box below.

Address: _____

Email: _____

- Receiving individual feedback after all data has been collected.
YES NO

.....
Signature

.....
Please PRINT name

.....
Participant #

.....
Date

Appendix E – Continued

Physical Activity Readiness Questionnaire

©Canadian Society for Exercise Physiology

Regular physical activity is fun and healthy, and increasingly more people are starting to become more active every day. Being more active is very safe for most people. However, some people should check with their doctor before they start becoming much more physically active.

Common sense is your best guide when you answer these questions. Please read the questions carefully and answer each one honestly: check YES or NO.

Yes No

- Has your doctor ever said that you have a heart condition and that you should only do physical activity recommended by a doctor?
- Do you feel pain in your chest when you do physical activity?
- In the past month, have you had chest pain when you were not doing physical activity?
- Do you lose your balance because of dizziness or do you ever lose consciousness?
- Do you have a bone or joint problem (for example, back, knee or hip) that could be made worse by a change in your physical activity?
- Is your doctor currently prescribing drugs (for example, water pills) for your blood pressure or heart condition?
- Do you know of any other reason why you should not do physical activity?
- Are you pregnant?

If you answered:

Yes; talk to your doctor before becoming physically active

No; you can be reasonably sure that you can start becoming physically active, and can participate in our study.

Printed Name _____

Participant # _____

Signature _____

Date _____

Appendix F – Pre-Study Questionnaire

Pre-Test Participant Questionnaire

Participant #: _____

Q1. At the moment, what sporting (competitive) activity are you involved in?

Sporting Activity	Hours per week	Sporting Activity	Hours per week

Q2. Prior to this year, can you report your sporting activity experience?

Sporting Activity (e.g., swimming)	Years involved (e.g., 2-3 years)	Hours per week (e.g., 8 hours per week)	Level of Participation (e.g., county level)

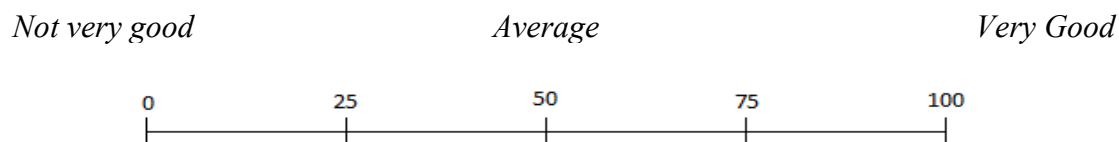
Appendix G – Pre and Post Task Questionnaires

Pre/Post Task Questionnaire – 60m Sprint

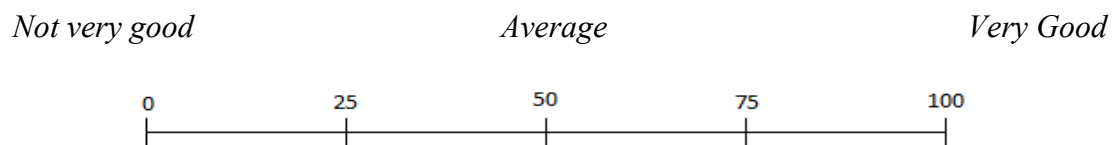
Please use the following scale standard scale to indicate your responses. Try to estimate as precisely and accurately as you can (e.g., 66%).

Pre-Task

Q1. Compared to the average student at your university, where would you rank your performance time on a 60m sprint task?

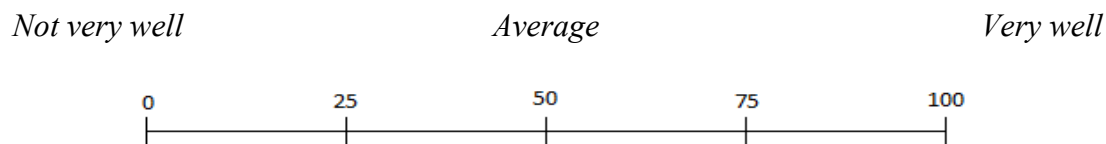


Q2. Compared to the average athlete your age (+/- 3yrs) outside your university, where would you rank your performance time on a 60m sprint task?

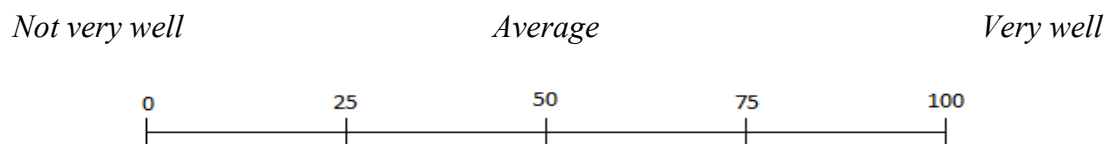


Post-Task

Q3. Compared to the average student at your university, where would you rank your performance time on the 60m sprint task?



Q4. Compared to the average athlete your age (+/- 3yrs) outside your university, where would you rank your performance on the 60m sprint task?



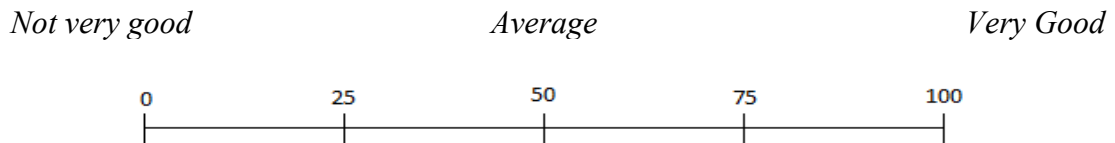
Appendix G – Continued

Pre/Post Task Questionnaire – Vertical Jump

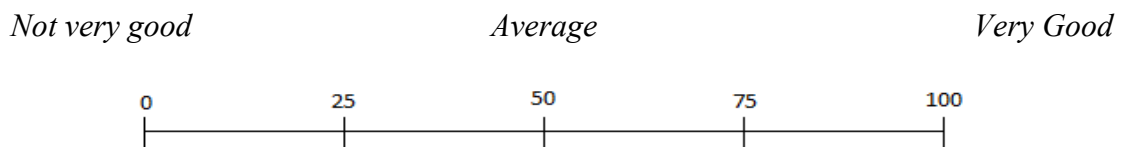
Please use the following scale standard scale to indicate your responses. Try to estimate as precisely and accurately as you can (e.g., 66%).

Pre-Task

Q1. Compared to the average student at your university, where would you rank your performance height on a vertical jump task?

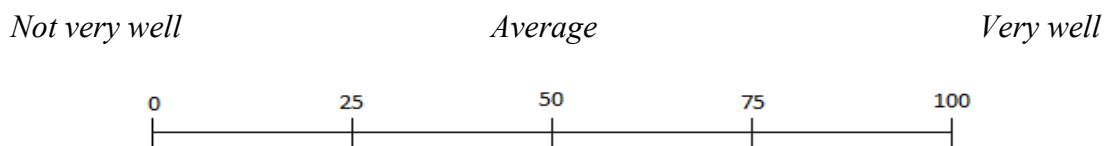


Q2. Compared to the average athlete your age (+/- 3yrs) outside your university, where would you rank your performance height on a vertical jump task?

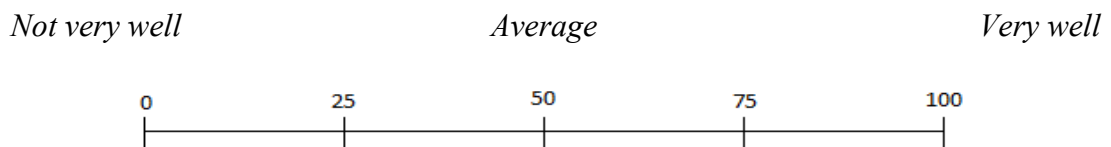


Post-Task

Q3. Compared to the average student at your university, where would you rank your performance height in the vertical jump task?



Q4. Compared to the average athlete your age (+/- 3yrs) outside your university, where would you rank your performance height in the vertical jump task?



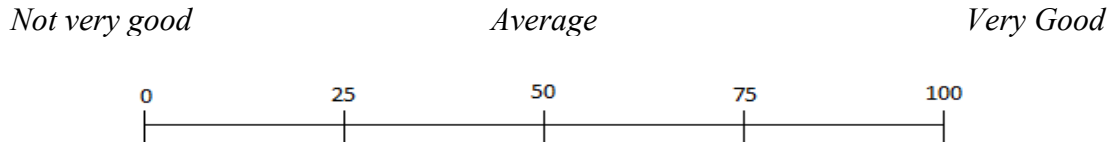
Appendix G – Continued

Pre/Post Task Questionnaire – Stroop

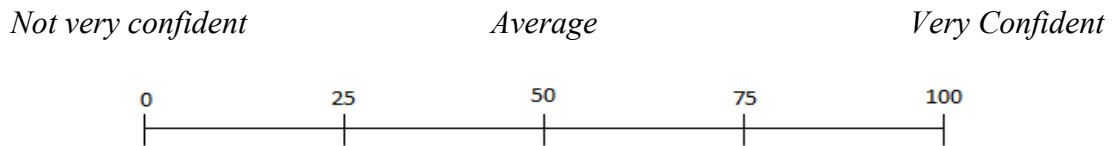
Please use the following scale standard scale to indicate your responses. Try to estimate as precisely and accurately as you can (e.g., 66%).

Pre-Task

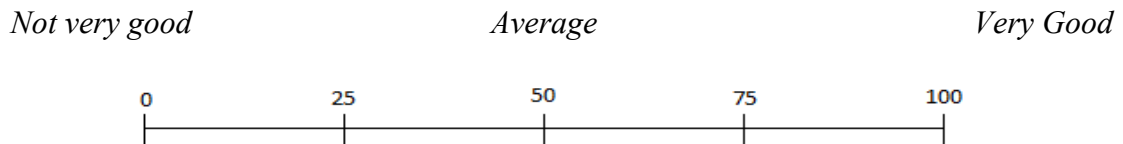
Q1. Compared to the average student at your university, where would you rank your performance time in a response time (Stroop) task?



Q2. Compared to the average student at your university, how confident are you in being able to perform in the reaction time task close/near to your personal best?



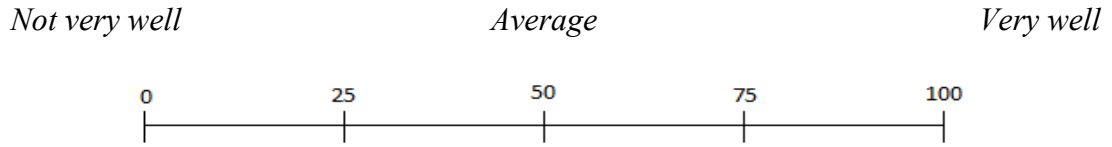
Q3. Compared to the average athlete your age (+/- 3yrs) outside your university, where would you rank your performance time in a response time (Stroop) task?



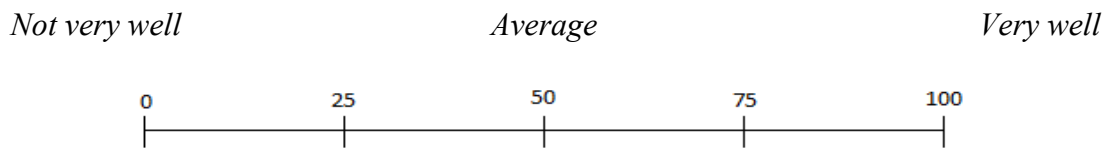
Appendix G – Continued

Post-Task

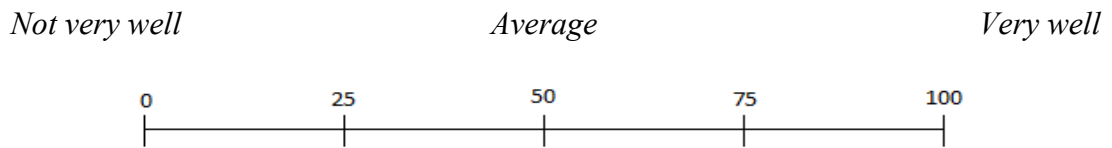
Q4. Compared to the average student at your university, how well do you think you have performed in the response time (Stroop) task?



Q5. Compared to the average student at your university, how confident are you that you have performed the task close/near to your personal best?



Q6. Compared to the average athlete your age (+/- 3yrs) outside your university, where would you rank your performance in the response time (Stroop) task?



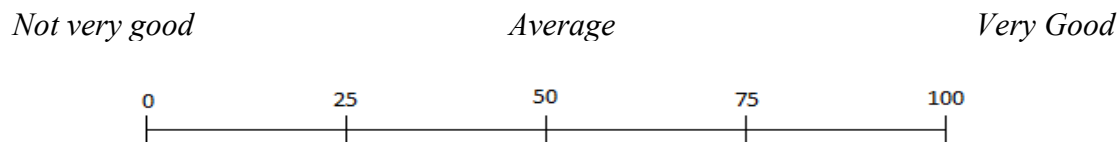
Appendix G – Continued

Pre/Post Task Questionnaire – Tower of Hanoi

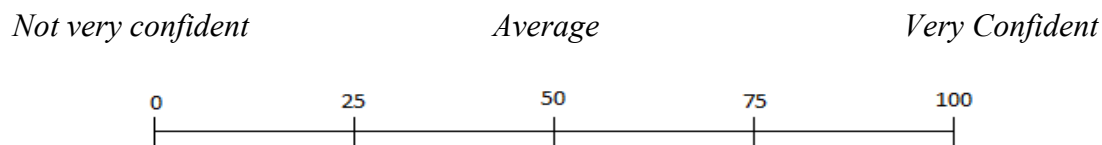
Please use the following scale standard scale to indicate your responses. Try to estimate as precisely and accurately as you can (e.g., 66%).

Pre-Task

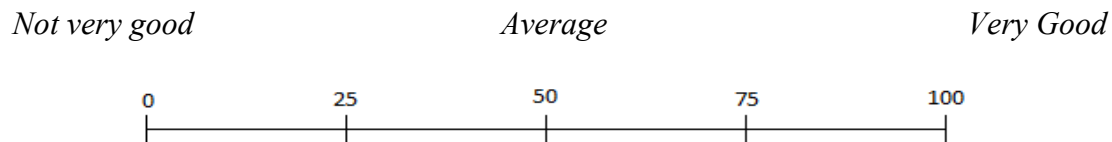
Q1. Compared to the average student at your university, where would you rank your performance time on a Tower of Hanoi task?



Q2. Compared to the average student at your university, how confident are you in being able to complete the task close/near to your personal best?



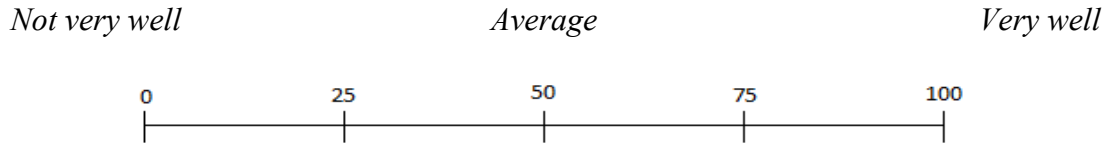
Q3. Compared to the average athlete your age (+/- 3yrs) outside your university, where would you rank your performance time on a Tower of Hanoi task?



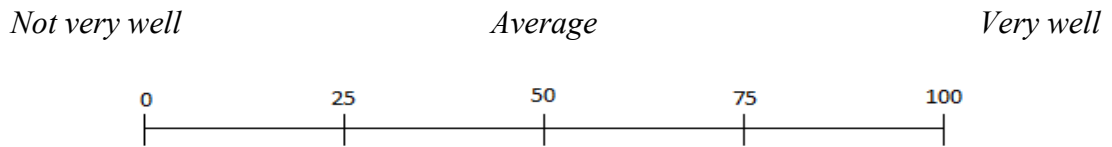
Appendix G – Continued

Post-Task

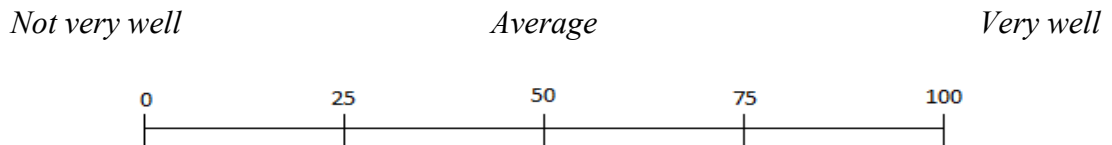
Q4. Compared to the average student at your university, where would you rank your performance time in the Tower of Hanoi task?



Q5. Compared to the average student at your university, how confident are you that you have completed the task close/near your personal best?



Q6. Compared to the average athlete your age (+/- 3yrs) outside your university, where would you rank your performance time in the Tower of Hanoi task?



Appendix H – Data Collection Sheet

Results Sheet

Participant #

Height..... Weight..... Handedness R / L

1 60 Meter Task:

- Warm-up min

Trial 1 30mseconds 60mseconds

Trial 2 30mseconds 60mseconds

Trial 2 30mseconds 60mseconds

2 Tower of Hanoi Task

Time to completion (minutes/seconds), max 10 min

Moves Completed Y / N

Rings left unmoved Tallest Tower

3 Vertical Jump Task

- Practice jump (One)

- Standing Heightcm

Jump attempt 1cm

Jump attempt 2cm

Jump attempt 3cm

4 Reaction Time

Results recorded on computer.

Appendix I – Participant Debrief

The Dunning-Kruger Effect as a Constraint on Learning and Expertise: Insights from Motor Performance Contexts

AKA

Psychological Processes in Sport

PARTICIPANT DEBRIEF

We would like to thank you for participating in this experiment. Regardless of how you think you may have done, the information you have provided us will be of great value.

The experiment you were involved with was looking at the differences in your actual versus perceived performance in the tasks that we set for you. As seen in other domains, the differences between your actual and perceived performance when compared to your peers is called the Dunning-Kruger Effect. We were investigating whether that effect exists in motor performance contexts. We were also looking at the potential psychological mechanisms that might explain why people show this effect.

The Dunning-Kruger Effect is when people who perform poorly, overestimate their performance when comparing to their peers; and people who perform very well, underestimate their performance when comparing to their peers.

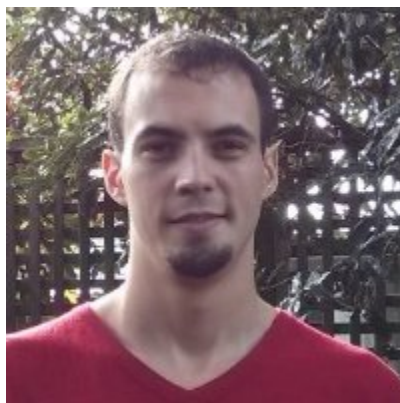
Psychological mechanisms that we looked at for being potential causes of this effect are your: level of procrastination, pessimism, optimism, self-presentation tactics usage, and orientation of motivation,

Our overarching aim is to see how this effect constrains persons from learning skills in motor performance contexts. Results from this study could have significance in numerous fields such as education, sport, business and the military.

Currently data collection has not been completed so we are unable to give you the results of this study, however if you are interested please contact us and we will be able to send you a link to the final paper once published.

Appendix J – Summary of Masters Candidature Contribution

Tate Hubka



Proposed thesis publications:

Hubka, T., Copley, S., & Adams, R. (2014). The capability to match estimated performance with actual performance: A meta-analysis. *Manuscript submitted for publication.*

Hubka, T., Copley, S., Adams, R., & O'Connor, D. (2014). The capability to align estimated and actual performance: The Dunning-Kruger effect is evident in physical task contexts. *Manuscript submitted for publication.*

Hubka, T., Copley, S., & Adams, R. (2014). The Dunning-Kruger Effect in Cognitive Performance Contexts. *Manuscript in preparation.*

Proposed Masters publications:

Brightmore A., O'Hara, J., Till, K., Copley S., Hubka, T., Emmonds, S., & Cooke, C. (2014). Comparison of movement demands in Rugby League referees between the European Super League and Australasian National Rugby League. *Manuscript in preparation.*

Hubka, T., Copley, S., Adams, R. (2014). (In)Accurate capabilities to align estimated and actual performance: The Dunning-Kruger effect in motor performance contexts. *Manuscript in preparation.*

Soomro, N., Copley, S., Hackett, D., Freeston, J., Hubka, T., Ibrahimi, S., & Sanders, R. (2014). Meta-analysis on efficacy of injury prevention programs in adolescent sports. *Manuscript in preparation.*