FORMING JUDGMENTS AND THE INFLUENCE OF GROUPING ITEMS

by

Lindsay Valve

A thesis submitted in conformity with the requirements for the degree of Masters of Arts in Education
Graduate Department of Curriculum, Teaching, and Learning
Ontario Institute for Studies in Education
University of Toronto

© Copyright by Lindsay Valve (2015)
FORMING JUDGMENTS AND THE INFLUENCE OF GROUPING ITEMS
Masters of Arts in Education (2015)
Lindsay Valve
Department of Curriculum, Teaching and Learning
University of Toronto

Abstract
This study examined the judgment strategies used by doctoral students to self-assess their research skills, and the effects of questionnaire item grouping (thematic versus random) on the selection of those strategies. Self-identification of researcher type (quantitative versus qualitative) was also evaluated. A cognitive interview approach using “think aloud” and retrospective probing protocols were used with a two-group (control/random and item-grouped) design. Findings revealed participants used one main and at least one supporting judgment strategy to formulate their self-assessment. The authors found no evidence that item grouping affected strategy selection. Item grouping did seem to influence assessment accuracy on quantitative research skills, mediated by participants’ researcher type (qualitative or quantitative researcher)
Table of Contents

Abstract.............................................................................................................................. ii
Table of contents .................................................................................................................. iii
List of tables ......................................................................................................................... v
List of figures ....................................................................................................................... vi

Chapter 1: Introduction ........................................................................................................... 1
  Research Questions ............................................................................................................. 9

Chapter 2: Literature Review ................................................................................................. 10
  Cognitive Processes ........................................................................................................... 11
    Forming judgments ......................................................................................................... 12
    Reporting on attitudes .................................................................................................. 14
    Reporting on behaviours .............................................................................................. 15
    Reporting accuracy ....................................................................................................... 15
    Cognitive interviewing ................................................................................................. 17
    Concurrent protocols: think-aloud and probing ............................................................ 20
  Item Grouping .................................................................................................................. 24
    Random and systematic error ....................................................................................... 24
    Construct validity of self-assessment .......................................................................... 25
    Context effects .............................................................................................................. 26
    Item-priming effects ...................................................................................................... 26
    Item-grouping effects ................................................................................................. 27

Chapter 3: Methods ............................................................................................................... 32
  Participants ....................................................................................................................... 32
  Protocol ............................................................................................................................ 34
    Self-assessment instruments ......................................................................................... 35

Data Collection and Analysis .............................................................................................. 36

Chapter 4: Results ................................................................................................................. 37
  Participant Characteristics ............................................................................................... 37
    Year of study .................................................................................................................. 38
    Research experience – as researchers ....................................................................... 38
    Research experience – as participants ....................................................................... 39
    Defining “experience” ................................................................................................. 39
    Cognitive strategies ...................................................................................................... 40
    Referring to a specific event ....................................................................................... 41
    Referring to a general pattern regarding the target attribute ..................................... 42
    Referring to other’s perceptions .................................................................................. 43
    Referring to a related general trait ............................................................................. 43
    Referring to a consequence of the attribute ............................................................... 44
    Breaking the attribute down into specific elements and evaluating each element .... 45
    Referring to an associated feeling .......................................................................... 46
    Relying on a socially desirable response .................................................................. 47
    Qualitative vs. quantitative ......................................................................................... 48
  Item-Grouping .................................................................................................................. 50

Chapter 5: Discussion ........................................................................................................... 53
List of Tables

Table 1: Participant demographic information ................................................................. 33
Table 2: Major judgment strategies .................................................................................. 38
Table 3: Summary statistics by question for non-grouped and item-grouped conditions ... 41
Table 4: Summary of main and secondary judgment strategies employed by participants 43
Table 5: Summary of qualitative and quantitative means by treatment condition .......... 53
List of Figures

Figure 1: Mean Qualitative, Quantitative, and Total Scores by Treatment Condition ....... 53
Chapter 1: Introduction

Return-on-investment is a priority for most organizations today. Every decision made must be justified; and the data points used and weighed to evaluate the return-on-investment are numerous. A major area of interest for many organizations is training evaluation. The appeal of developing “learning organizations”, and the popularity of placing “people development” at the forefront of organizational values has increased considerably over the last 20 years, and continues to rise (Garaway, 1995; Fitzpatrick et. al., 2011).

Understanding the performance (impact) of a training intervention and evaluating the post-training performance of respondents (outcomes) are priorities for most. It is this data that provides assurance that the budget invested in training initiatives will translate into improved business outcomes (Bates, 2004). Without this information, little can be said about the value of an intervention.

Organizations employ a variety of strategies to achieve this goal, ranging from the implementation of comprehensive measurement strategies (such as performance measurement strategies; Mayne, 1999), to less rigorous approaches (such as post-test-only evaluations; Bamberger, et. al, 2004) to estimate training effects. Customary in many of these approaches is the use of self-assessment (e.g., Marsh & Roche, 1997; Harwell, 1999; Moran & Hoy, 2001; Collie-Akers, Watson-Thompson, Schultz & Fawcett, 2010). In this context, self-assessment, in which individuals are asked to judge themselves in response to a questionnaire or interview, is often used to assess training outcomes and impacts (e.g., Sadri & Snyder, 1995; Alvarez, Salas & Garofano, 2004; Attia & Honeycutt, 2012).
Naturally, in the case of self-assessment, asking respondents to evaluate and form judgments of themselves introduces a number of potential biases. These biases include intentional biases, when a participant’s goals or personal motivations consciously influence the way they make judgments (e.g., self-serving bias); and unintentional biases, which occur sub-consciously due in part to lack of experience in self-assessment (Murphy et al., 1989; Leckie & Baird, 2011). Biases in judgment are mediated by a variety of factors, including what is being assessed (i.e., the specific behaviour(s), skill(s), belief(s), attitude(s)), the respondent’s emotional investment in their progress, and the relationship between the training content and the respondent’s self-concept (Pratt, McGuigan & Katzev, 2000; Taylor, Russ & Taylor, 2009). Ultimately, biases in participants’ judgments pose threats to the validity of the self-assessment, and bring into question the decisions made using this data (Pratt, McGuigan & Katzev, 2000; Bates, 2004; Taylor, Russ & Taylor, 2009).

Strategies to mitigate these errors in self-assessment have been recommended, and vary based on the specific evaluation scenario (Lam & Bengo, 2003; Leach & Liu, 2003; Lam, Kolomitro & Alamparambil, 2011). The majority of these alternatives address procedural flaws in the traditional self-assessment (e.g., Dwyer, 1980; Kruger & Dunning, 1999; Fitzgerald, White, & Gruppen, 2003; Stufflebeam & Wingate, 2005), respondent’s lack of knowledge or skill at pre-test time (e.g., Ehrlinger, 2007), and situational variables related to the motivations, confidence, and personal stakes of the respondents. In other words, current methods attempt to manage a participant’s internal processes by managing external variables. Unfortunately, despite these efforts, there has been little improvement to the validity of self-assessment data, as these alternative approaches tend to introduce new biases. For example, the retrospective post-pretest was developed to prevent response-shift bias (a change in the pre and post-test measure resulting
from respondent’s improved knowledge at the post-test); however, this method also invited new issues related to biases, such as augmenting socially desirable responses and consequently the likelihood of overestimation in self-assessment, which compromise the assessments’ validity (Hill & Betz, 2005; Moore & Tananis, 2009; Nimon, Zigarmi & Allen, 2011). Other strategies for mitigating self-assessment biases have included rater training and skill refinement, assuming raters do not have the skills to self-assess (Murphy, Cleveland, Skattebo & Kinney, 2004); asking raters to compare themselves to other, peer, participants to “level set” the benchmark against which they are self-assessing (Dunning, Meyerowitz & Holzberg, 1989); and giving raters feedback about their strengths and weaknesses to prevent overestimation (Dunning, Heath & Suls, 2004).

The persistence of significant biases in self-assessment suggests that the strategies developed to-date do not address their primary source. Interestingly, there is a notable absence of approaches attending to the cognitive strategies respondents use to form judgments (i.e., that enable self-assessment), and their mediating factors. An examination of these cognitive strategies would consider the steps respondents take to generate a self-assessment, which might provide further insight into when and how biases are introduced. For example, such an examination might explore the information respondents take into consideration, and how they use this information to generate a self-assessment. Further deconstruction of these cognitive steps might provide insight into when and how information is used inaccurately, or when flawed information is referenced, leading to over/underestimations and other biases. While a preliminary understanding of judgment strategies is available, deeper insight into these processes could be used to develop self-assessment methodologies that target specific areas of cognitive failure in order to mitigate the introduction of certain biases. In other words, a more robust understanding
of respondent’s judgment strategies would facilitate an evaluation of the cognitive (i.e., internal) processes that occur in self-assessment.

Research examining the cognitive strategies respondents employ during a self-assessment is scarce; although the examination of cognitive strategies in other contexts has proven to be an effective approach to uncovering the way respondents comprehend, interpret, and act upon questionnaire items (e.g., Tourangeau & Rasinski, 1988; Hogarth & Einhorn, 1992; Melnick, 1993; Lam, Green & Bordignon, 2002). This work has revealed a set of four, general cognitive phases respondents move through to generate evaluative responses about their experiences: comprehension (forming an internal representation of the question, and serving as a starting point for information/memory retrieval); retrieval (the process of searching memory for the information requested, either through direct recall or using an estimation heuristic that might deal with partially-recalled/inadequate information); judgment (evaluation of information that guides the search for information and ultimately the output of information); and response (selection between response alternatives and output of information to be reported) (Tourangeau, 1984; Willis, 2005). These four phases illustrate the fact that the retrieval of information (generation of a response) is not automatic; rather, that there are judgment processes involved both before and after information is retrieved and formulated into a response (Tourangeau, 1984; Willis, 2005; Kahneman, 2011). The judgment strategies used to guide and evaluate the information retrieved have also been examined, to an extent (Lam, 2014). These strategies include: (1) referring to a specific event (Krosnick, 1991), (2) thinking about a general pattern involving the target attribute (e.g., considering how they tend to manage school work and perform in school), (3) referring to others’ perceptions (Conrad, Brown & Cashman, 1998), (4) reflecting on a related general trait (Heine & Lehman, 1995), (5) referring to a consequence of the attribute being measured (Heine
& Lehman, 1995), (6) referring to the feeling(s) associated with the attribute (Kelley & Lindsay, 1993; McKenzie, 1994), (7) breaking the target attribute down into specific elements (operationalizing) and evaluating each one (Krosnick, 1991), and (8) relying on a socially desirable response (Krosnick, 1991), or other self-serving purposes (Conrad, Brown & Cashman, 1998).

Consistent with self-assessment research that highlights validity threats resulting from respondent’s understanding, motivation, and other biases, the mental steps used to generate judgments are vulnerable to both intentional (e.g., Tourangeau & Rasinski, 1988) and unintentional errors (e.g., Lam, Green & Bordignon, 2002). These vulnerabilities translate into specific types of failure that can occur at each phase of cognition (comprehension, retrieval, judgment, response), and have varying effects on responses.

Factors external to the respondent have also been demonstrated to have significant effects on the way judgments are formed. The item order effect, which refers to the effect of the response to one or more items, on the response to subsequent items (Tourangeau & Rasinski, 1988; Hogarth & Einhorn, 1992; Lam, Green & Bordignon, 2002), has been shown to impact the ways in which items are processed; causing an interaction between the information processing strategies used by respondents and specific task characteristics (Tourangeau & Rasinski, 1988; Hogarth & Einhorn, 1992; Lam, Green & Bordignon, 2002; Ortner, 2008; Marentette, Meyers, Hurtz & Kuang, 2012). For example, Tourangeau and Rasinski discussed that prior items on an assessment can affect participants’ judgment formation by changing the anchor for the response scale (1988).
Item grouping effect, which is a special form of item order effect, is of particular relevance to self-assessment. Item grouping examines the effects of grouping items (e.g., by theme, construct, sub-scale) together on a questionnaire or during an interview (Melnick, 1993; Lam, Green & Bordignon, 2002). There are two types of item grouping: with labels (item groups are clearly labeled and might include a definition of the construct preceding the assessment items), and without labels (items are grouped but these groupings are not made explicit through labels or other indicators). In both scenarios, the underlying premise is that by grouping items, respondents are provided with context and a consistent point of reference against which to base their judgments.

Research in this area has generated conflicting points of view, with some studies concluding “no effect”, while others report a positive effect on both the respondent (e.g., by facilitating the mental processing involved) and improving the internal reliability of the measures (Tourangeau, 1984; Tourangeau & Rasinski, 1988; Melnick, 1993; Lam, Green & Bordignon, 2002). Other research suggests that item grouping causes a negative effect by inviting a generalized response to a group, causing participants to be inattentive to the individual items within each group (Strack, Martin & Schwarz, 1988; Ortner, 2008; Bowman & Schuldt, 2014).

It is clear that self-assessment is widely vulnerable to a number of independent and combined effects, all of which pose validity threats and compromise the integrity of the results. The relative ineffectiveness of alternative self-assessment methods established to mitigate these threats, combined with a poor understanding of the judgment strategies and moderating factors involved in self-assessment, and the variety of respondent-related and instrument-related factors affecting self-assessment accuracy, point to judgment strategies as a reasonable source of potential error warranting further examination. While an abundance of studies have conducted
quantitative analyses of response patterns, and made inferences about the cognitive strategies used to generate those patterns (Tourangeau, 1984; Willis, 2005), a qualitative examination of these strategies for making judgments and the influence of item grouping on responses and respondent’s cognitive processes is lacking.

Given the private nature of respondent’s internal thoughts, the quantitative approach taken to-date, in lieu of examining these thoughts, is understandable. In order to gain access to these internal thoughts, respondents would need verbalize (externalize) their thought processes and judgment strategies, while attempting to limit biases in what and how much is shared. Cognitive interviewing, the process of inviting respondents to vocalize their thinking during an assessment, was designed to expose the normally internal decision-making process, and reveal the cognitive steps the respondents follow to arrive at their answers (Tourangeau, 1984; Willis, 2005). Exposure to participants’ judgment processes could reveal the point(s) at which both intentional and unintentional biases being to influence cognitive processing, specifically the aforementioned four phases. The free-flowing narrative encouraged through the “think aloud” process of cognitive interviewing (in which participants are asked to verbalize every thought as it comes to mind) might also shed light on the sources of these biases, as respondents choose between examples, weigh information, and ultimately articulate a judgment; while follow-up probing would provide respondents with the opportunity to expand upon the rationale for their thought processes and judgment strategies.

Although cognitive interviewing has been used predominantly to pilot test questionnaire items to determine the degree to which respondents interpret items in ways as intended by the researchers, its application to the confirmation of construct validity (i.e., by ensuring the cognitive strategies being applied by the respondent are suitable for the construct being
measured) has increased in recent years (Schwarz, 2007; Brown, Hawkes & Tata, 2009; Andersen, Groenvold, Jorgensen & Aadhl, 2010; AERA/APA/NCME, 2014). A similar approach might be taken to examine the cognitive strategies employed by respondents during a self-assessment.

The use of self-assessment is pervasive across industries, accompanied by errors endemic to the methodology; and the potential consequences of using error-prone data to draw conclusions and form the basis of decisions, are substantial. Alternative strategies currently overlook the role of cognitive strategies in the self-assessment process, despite related research that upholds the vulnerability of a respondent’s cognitive processing to context and other effects (such as item order and item grouping). The purpose of this study is therefore two fold. First, it is to examine the cognitive strategies respondents use to form judgments during a self-assessment; and second, to examine the effect of item grouping on item responses, judgment formation and the selection of cognitive strategies.

The omnipresence of self-assessment use in training evaluation across industries forms the basis of the rationale for this study. Research to-date has been limited in its focus, prioritizing almost exclusively the quantitative analysis of responses; and inferences about the mechanisms involved in generating those responses. This study would contribute to an in-depth understanding of self-assessment by exploring the cognitive processes involved. This is an area not currently addressed in the research, and would provide valuable insight into the ways a respondent forms judgments, including the internal and external factors that influence those decisions. Understanding these judgment strategies, including when and how biases are introduced during the judgment process, might be used to improve construct validity; and to inform item and scale construction.
Self-assessment methodologies tend to adhere to predictable questionnaire and survey formats. Given the self-directedness of this type of assessment, respondents are often left to their own devices to interpret and respond to question items. This potential ambiguity renders them more susceptible to moderating factors like item order and grouping effects. By explicating the respondent’s decision-process, whether and how these factors (such as item grouping) influence the specific strategies a respondent employs will be revealed. This insight might be used to inform improved questionnaire and interview design, potentially increasing the validity of self-assessment responses.

The purpose of this research is to explore the strategies participants use to form judgments about themselves, the effects of grouping items in a survey on the responses and validity of these responses.

Research Questions

1. What cognitive strategies do respondents use to evaluate the specific aspects of their research efficacy?
2. How does item grouping affect respondents’ judgment forming strategies?
3. How does item grouping affect respondents’ responses and validity?
Chapter 2: Literature Review

Self-assessment, in which individuals are asked to analyze or judge themselves by responding to a questionnaire or interview, is used throughout training evaluation and in educational research and evaluation in general. Research from numerous areas in psychology, assessment, and education suggest that self-assessments are often flawed in substantive and systematic ways, which introduce errors and consequently threats to the validity of the data generated (Cook & Campbell, 1979; Lam & Bengo, 2003; Bamberger et al, 2004; Dunning, Heath & Suls, 2004; Nimon, Zigarmi & Allen, 2011). These flaws are arguably inherent in the act of judging oneself, as it considers a single perspective vulnerable to personal motives and biases.

Research on the validity of self-reported data has revealed a number of mediating factors, including individual (respondent) differences, instrument factors, contextual influences, assessment content, and procedural factors (Lam & Bengo, 2003; Bamberger et al, 2004; Lam, 2009; Nimon, Zigarmi & Allen, 2011). For example, more capable learners tend to underrate their performance, while less-able learners tend to overrate their performance (Lam, 2009); high intelligence, high achievement status, and internal locus of control are all associated with improved self-assessment accuracy (Mabe & West, 1982); self-assessment of performance is more accurate after the performance than before (Stufflebeam & Wingate, 2005); and measurement conditions such as anonymity and using criterion measures tend to improve accuracy (Stufflebeam & Wingate, 2005).

Self-assessment methodologies have been created in an effort to mitigate these known factors; however, given the breadth of potential risk factors, methods that target one issue (e.g., instrument or procedural factors) often increase the assessment’s vulnerability to other threats
(e.g., individual differences). In order to provide a more complete understanding of this challenge, the following examines the major factors moderating self-assessment accuracy, including mediating variables and bias effects, in more detail.

**Cognitive Processes**

Cognitive processes are the basic mental activities respondents use to think, to interpret incoming information, to recall information from memory, to judge information, and to formulate responses (Schwarz, 2007). During a self assessment, these strategies are leveraged within the context of four distinct cognitive phases: (1) comprehension (forming an internal representation of the question, and serving as a starting point for information/memory retrieval), (2) retrieval (the process of searching memory for the information requested, either through direct recall or using an estimation heuristic that might deal with partially-recalled/inadequate information), (3) judgment (evaluation of information that guides the search for information and ultimately the output of information), and (4) response (selection between response alternatives and output of information to be reported) (Tourangeau, 1984). Intuitively, these phases progress linearly as respondents are prompted to begin retrieving and judging encoded information upon the presentation of a question. Willis clarified that the sequence is still flexible, however; noting that respondents might cycle between retrieval and judgment a number of times in order to generate what they perceive to be an appropriate response (2005).

The presence of these cognitive phases illustrates that the retrieval of information is not automatic; rather, that there are judgment processes controlling the retrieval of information and ultimately the response (Willis, 2005). This conscious control renders cognitive strategies vulnerable to both intentional (deliberate) and unintentional sources of error (bias), meaning that failure is possible at each phase of cognition.
Forming Judgments. Self-assessments require respondents to form judgments about attributes such as their knowledge, skills, attitudes, and affects (including efficacy and motivation). The formation of these judgments requires respondents to reflect on personal experiences within the construct being measured, and to evaluate their memories in order to generate a response (Krosnick, 1991). In other words, the process of self-assessment relies on respondent’s ability to accurately move through the four phases of cognitive processing by making a series of decisions about the relevance and appropriateness of their memories, and to generate a view of their performance or efficacy within the construct being measured.

The processes underlying decision-making have been widely researched; with dual process theories among the most enduring explanations for the way respondents process these types of tasks (Stanovich & West, 2000). While the technical properties of these theories differ to varying degrees, the prototypical view describes two systems for processing judgments (Stanovich & West, 2000). System 1 is described as automatic, emotional, intuitive, and cognitively undemanding; while System 2 is considered a model of controlled processing. It is deliberate, calculating, analytical, and logical. The most important difference between these two systems has been described as differentiated task construal (Stanovich & West, 2000). That is, each system perceives the purpose of the same task differently, ultimately cueing different participant responses.

The System 1 and System 2 processes are believed to occur sequentially, when triggered. Through a series of experiments, Kahneman (2011) demonstrated that all participants first generate a response through System 1, and that with effort; participants might subsequently take a more linear approach to decision-making by engaging System 2. These results might be expanded upon to link to the earlier discussion of Tourangeau’s cognitive phases. Combining
these two areas of research, it appears that the system of thought engaged would determine the effort applied in the retrieval and judgment phases. The comprehension phase, therefore, might provide the information used to determine which system to rely on (i.e., whether to rouse System 2 from its baseline state).

Judgment formation is not automatic once System 1 or 2 has been selected. Judgment strategies refer to the specific approaches respondents take to move through the four phases of cognition in order to form judgment. In the process of engaging Systems 1 and/or 2, respondents employ judgment strategies to guide the system’s retrieval and evaluation of information. Literature on self-assessment, decision-making, and memory has revealed that respondents use a variety of strategies to form judgments. Most notably, respondents tend to rely on a variety of estimation strategies (e.g., Gilbert, Krull & Pelham, 1988; Blair & Burton, 1987; Griffin & Tversky, 1992; Kelley & Lindsay, 1993). Lam (2014) noted eight cognitive strategies for making judgments. The following is the listing of these strategies and references to publications that have also noted each of these strategies: (1) referring to a specific event (Krosnick, 1991), (2) thinking about a general pattern involving the target attribute (e.g., considering how they tend to manage school work and perform in school), (3) referring to others’ perceptions (Conrad, Brown & Cashman, 1998), (4) reflecting on a related general trait (Heine & Lehman, 1995), (5) referring to a consequence of the attribute being measured (Heine & Lehman, 1995), (6) referring to the feeling(s) associated with the attribute (Kelley & Lindsay, 1993; McKenzie, 1994), (7) breaking the target attribute down into specific elements (operationalizing) and evaluating each one (Krosnick, 1991), and (8) relying on a socially desirable response (Krosnick, 1991), or other self-serving purposes (Conrad, Brown & Cashman, 1998).
The interplay between Systems 1 and 2, and judgment strategy selection are still being investigated, and show a fluid relationship guided by a variety of factors and variables. For the purposes of this research, two main factors influencing a respondent’s approach to forming judgments will be considered: the types of attributes respondents are typically asked to evaluate or report on; and the conditions in which they are asked to report. That is, self-assessment content and context should be considered mediating variables of judgment strategy selection, and respondent’s overall approach to forming judgments.

**Reporting on Attitudes.** There are two prevailing perspectives regarding the permanence of attitudes, which result in differing opinions regarding respondent’s abilities to form judgments about those attitudes. Many psychologists conceptualize attitudes as enduring dispositions, and contextual influences on respondent’s attitudes as a source of distraction that clouds their “true” attitudes (Schwarz, 2007). Alternatively, attitudes are also treated as evaluative judgments formed on the spot, based on the information available in the moment (Schwarz, 2007). From this perspective, contextual influences on attitude judgments are part of the phenomenon of interest, rather than artificial (Schwarz, 2007). In fact, it is expected that evaluations are informed by past and current experiences; memories of both similar and dissimilar situations; and should even take into consideration the respondent’s current goals, weighing each of these sources of information in accordance with the judgment task (Schwarz, 2007). This latter perspective is upheld by numerous research observations that attitudes are highly context dependent (Schuman & Presser, 1981; Tourangeau & Rasinski, 1988), which reinforces the context and individual factors discussed earlier in this paper, and their influence on cognitive processing.
**Reporting on Behaviours.** The majority of what is known about human behaviour has been derived from self-reports collected in representative sample surveys (Schwarz, 2007). As discussed in relation to self-assessment, the accuracy of these self-reports have long been scrutinized given respondent’s propensity to misunderstand what they are being asked to report on; to forget information, or to filter and edit their responses (Schwarz, 2007). As research has improved understanding of the complexity of autobiographical memory, methods to facilitate recall have been developed (Schwarz, 2007). These methods tend to emphasize providing respondents with a context through which to interpret the question, and to guide their memory retrieval, judgment and responding (Schwarz, 2007). For example, Belli (1998) developed a method to facilitate the reporting of life history. This method provides respondents with a grid and asks them to place various behaviours and events in time and space, capitalizing on the observation that recalling information in one domain is likely to prime temporally or thematically related information from other domains (Schwarz, 2007). The culmination of these research reports confirms the context-dependency of judgments of behaviour (Schwarz, 2007).

**Reporting Accuracy.** In terms of reporting accuracy, a review of the literature clarifies that memory processes are only one factor (Conrad, Brown & Cashman, 1998; Schwarz, 2007). Intentional and unintentional biases in reporting undermine reporting accuracy, and can introduce a variety of threats to validity. For example, respondents may report on behaviours that do not match what the researcher is looking for (Schober & Conrad, 2002), contextual variables may influence question interpretation with downstream effects on judgment and reporting (Galesic & Tourangeau, 2007; Winkielman et al., 1998), and respondents might hesitate to report when they perform undesirable behaviours or fail to perform desirable ones (Schwarz, 2007). As with cognitive processing in general, these errors are exacerbated under conditions of ambiguity.

A source of judgment error that has been widely researched is the anchoring-and-adjusting heuristic, originally put forth by Tversky and Kahneman (1974). According to this theory, respondents begin by setting an anchor (e.g., an initial piece of information offered), and arrive at a final answer by adjusting away from the anchor (Tversky & Kahneman, 1974). The bias is introduced when respondents generally adjust insufficiently, resulting an answer that is closer to the anchor than it otherwise would be (there is a bias toward adjusting around the anchor). This strategy is of particular relevance in a self-assessment due to the influence of implicit theories, respondent mood and personality, and individual motivations on the setting of the anchor and their ability to adjust away from that anchor.

The variety of processes involved in forming judgments reviewed here underscores the complexity of self-assessment. Although a variety of approaches have been proposed to improve self-assessment and reporting accuracy, none take into consideration the complex interplay of cognitive phases, System 1 and System 2 judgments, the contextual-based attitude formation, intentional and unintentional biases, and the selection of judgment strategies. To an extent, this gap in self-assessment methodology might be explained by researcher’s limited understanding of the interrelationships that exist between and among the factors that influence forming judgments. While the current literature describes a series of hypotheses and observations, it does not describe causative associations between content, context, and cognition. Moreover, although a number of biases have been identified and cautioned against, awareness of when these biases are introduced during the judgment-forming process is missing. An improved understanding of the
cognitive processes and judgment strategies involved in judgment formation; and importantly, awareness of the factors mediating the selection and application of these strategies, is warranted.

**Cognitive Interviewing.** Traditional surveys and interviews focus on responses, without accessing the cognitive strategies respondents employ and the decision-making steps they take to generate those answers. Instead, these internal processes are typically inferred by looking at respondent’s answers. Cognitive interviewing was designed to expose the cognitive strategies and steps respondents take pertaining to the four phases of cognitive processing. Initially, this method was part of the Cognitive Aspects of Survey Methodology movement (CASM), which revealed that respondents give inaccurate and/or unreliable information because they do not understand the questions, cannot remember the required relevant information, use flawed judgment strategies, have trouble mapping their internal judgments to one of the response options, or edit their answers in a way that is misleading before reporting them (Touragneau, 2003; Willis, 2005). These challenges seem to result from the ways respondents are used to answering questions in everyday life, and the shortcuts they are accustomed to taking in order to reduce the cognitive burden imposed by interviews and other tasks (Touragneau, 2003; Willis, 2005).

Although cognitive interviewing has been used predominantly to test questionnaire items, it is being leveraged increasingly to validate the cognitive processes of respondents as a way to confirm construct validity (i.e., making sure the cognitive processes being used are suitable for the construct being measured). Verbal report techniques, in particular, focus on the study of specific cognitive issues (usually pertaining to a particular questionnaire).

There are two main cognitive interview methods. Concurrent “think-aloud”, during which a researcher reads each question aloud and records (via audiotape or videotape) the respondent’s
verbal think-aloud stream (or in some other way notes the processes the respondent uses in arriving at an answer) (Ericsson and Simon (1980). The researcher does not interrupt the respondent at any time; instead, the respondent is free to speak anything and everything that comes to mind as it pertains to generating a response. This technique rests on the assumption that verbal reports reflect actual cognitive processes.

Despite the relative simplicity of conducting a think-aloud interview, the analysis of the results of these interviews is often variable and complex (Willis, 2005). The verbal stream produced by thinking aloud is submitted to a protocol analysis (detailed coding and review of the verbal output), through which behaviours and/or patterns are identified (Willis, 2005). Acknowledging the potential for misanalysis, Ericsson and Simon (1980, 1984) identify two important premises that should guide the use of this technique: (1) Self-reports are veridical to the extent that they involve output from short-term memory (versus long-term memory); and (2) under this circumstance, respondents can provide self-reports through think-aloud interviews that are non-reactive (i.e., the act of thinking aloud does not contaminate the respondent’s verbal processes).

Verbal probing, the second main form of cognitive interviewing, involves the researcher asking the target question, and the respondent answering. The researcher then follows up (immediately or at the end of the interview) by probing for other information related to the question or to the answer given (Willis, 2005). This technique requires that research make three key assumptions: (1) Features of the survey questions (e.g., question length, wording, etc.) impact the way a respondent processes information at each cognitive phase (i.e., comprehension, retrieval, judgment, response); (2) verbal probing techniques can be applied to “probe” the processing system and access the various sub-processes (e.g., comprehension, retrieval, etc.); and
(3) respondent’s verbal responses can be used to make judgments about the operation of the sub-processes (Willis, 2005). Additionally, researchers necessarily assume that probing does not bias the question processing (i.e., it does not produce reactivity) (Willis, 2005). While none of these assumptions can be guaranteed or proven, a number of recommendations regarding probe construction and interviewer behaviour have been made over the years to increase the likelihood that these assumptions are satisfied (Willis, 2005). Among these recommendations is a general agreement that in order to avoid reactivity effects due to the cumulative effects of probing, retrospective probing is often preferred. This practice should be balanced with the need to ensure that probed information is fresh in respondent’s minds at the time of probing to avoid having respondents fabricate an explanation (Snijkers, 2002).

Willis (2005) advised that these methods are not mutually exclusive; in fact following either method naturally generates activities and subsequent responses that are germane to the other method. Intentionally taking a combined approach to cognitive interviewing is therefore possible, but with caution, as it would likely be particularly taxing on the respondent. When choosing between these techniques, Willis provides the following analysis of benefits and drawbacks: the think-aloud technique assures freedom from interviewer-imposed biased (because the interviewer contributes little more than the reading of the survey question), minimal interviewer training requirements (the interviewer need only read the survey questions), and an open-ended format (minimal guidance of the respondent’s verbalization of their thoughts enables the sharing of unanticipated information). In contrast to these advantages, think-aloud does require respondent training (because thinking aloud is an unusual activity) and tends to mean limited think-aloud proficiency on the part of the respondent. Consequently, this technique places increased burden on the respondent, and leaves the outputs susceptible to respondent’s
information processing biases (e.g., thinking aloud might cause them to apply more effort than simply answering the question). Additionally, there is a tendency for the respondent to stray from the task (via free association), and it can be difficult to know whether the respondent is capable of answering the question (think-aloud encourages respondents to simply “talk”, not necessarily answer the question).

Verbal probing presents a similar set of advantages and drawbacks. Unlike think-aloud, verbal probing allows the researcher to maintain control of the interview (which avoids irrelevant and non-productive discussion), assures an investigative focus (by allowing researchers to examine what may be thought but left unstated), and requires little preparation or training of the respondent (Willis, 2005). In contrast, probing does present an increased potential for reactivity (because the interjection of probes produces a situation that is not a meaningful analog to the traditional survey interview), and increases the potential for bias through the misuse of probes (Willis, 2005). To mitigate these risks, verbal probing requires careful training of interviewers, including thorough preparation and selection of non-leading probing techniques (Willis, 2005).

**Concurrent Protocols: Think-aloud and probing.** The selection of an appropriate cognitive interview protocol is highly situational. While verbal reports have been demonstrated to account for a respondent’s mental processing before (what they would think), during (what they are thinking), and after (what they did think), the validity and reliability of this data has come under scrutiny (Taylor & Dionne, 2000). Specifically, whether and under what conditions participants’ cognitive processes can be reported have been questioned (e.g., Ericsson & Simon, 1993; Nibset & Wilson, 1977; Presley & Afflerbach, 1995; Van Someren, Barnard & Sandberg, 1994).

In an effort to examine the types of cognitive processing and knowledge that can be
accessed through cognitive interviewing, Taylor and Dionne (2000) used a combined approach of concurrent verbal protocols (think-aloud) and retrospective debriefings (probing) to examine participant’s problem-solving strategies. Problem-solving processes are akin to decision-making (judgment forming) processes, in that they rely upon complex interactions between cognitive and metacognitive strategies, experiences, knowledge, attitudes, beliefs, and social factors (Heppner & Krauskopf, 1987; Schoenfeld, 1983; Silver & Marshall, 1990; Taylor & Dionne, 2000). Some of these factors are under the control of the participant; while others are sub-conscious in their operation. As with forming judgments, these strategies are typically difficult to observe because they are internal – they occur “inside the participant’s mind” (Taylor & Dionne, 2000). The combined strategy used by Taylor & Dionne provided sufficient opportunity to access both conscious and sub-conscious processes: the think-aloud responses reflected the content of the problem-solving process (e.g., goals set, information used, decisions made along the way, rationale for the solution), while responses to the retrospective verbal probes included more specific references to the strategies controlling the problem-solving process, elaboration of the think-aloud content, and insights into the conditions that prompted particular thoughts or responses (Taylor & Dionne, 2000).

For the purpose of examining the ways in which participants form judgments and select and employ judgment strategies, the use of two complementary verbal recall protocols during cognitive interviewing seems appropriate. As demonstrated by Taylor and Dionne’s work, this combined strategy could be effective in examining attitudes that are formed either “on the spot”, as well as those that are informed through a longer, more enduring process of experiences (Schwarz, 2007). For example, a think-aloud protocol might reveal the steps a participant follows to form an attitude judgment “on the spot”, including the evidence and information the
participant considers, their weighting of this evidence, the memories of experience they might
draw upon, and the specific judgment strategies they employed. A retrospective follow-up probe
might then be used to explore how their feelings might have influenced their judgment, why they
weighted certain information more heavily than other evidence, and any other conditions that
affected how they chose to proceed. Such probes could also be used to ask more directly about
the judgment strategies they employed (e.g., “what did you think about when…”).

A combined protocol approach to cognitive interviewing might also be effective in
exposing the use of System 1 and System 2 judgment processes. Although engagement of each
system is sub-conscious, participant’s responses are likely to reveal indicators of the system
guiding the judgment-forming process. For example, think-aloud responses could reveal rapid,
intuitive judgments with limited elaboration or rationale. Retrospective follow-up probes might
be used to clarify the examples or other references the participant leveraged to arrive at their
responses. In this example, System 1 thinking might have been relied upon if no further
explanation could be given (i.e., the participant’s judgment remained intuitive), while System 2
thinking might be revealed during more elaborate think-aloud responses. Moreover, a think-
aloud approach might uncover the point at which System 2 thinking is recruited, as participants
expand upon certain aspects of their thinking more than others (without probing).

Literature challenging the validity of data collected from cognitive interviewing questions
the assumption that respondents have access to their cognitions; and can be relied upon to
provide accurate direct verbal reports (Willis, 2005). In opposition to this challenge, Ericsson &
Simon (1980; 1984) cited numerous studies showing that subjects are able to report on
intermediate results of a series of mental operations, and that the act of reporting does not in
itself strongly influence concurrent or subsequent behaviour. Taylor and Dionne’s research
provides example in favour of the accessibility of participants’ own cognitive processes.

In order to maximize the reliability of the information generated through cognitive interviewing, previous studies have confirmed improved accuracy of responses when the interview involves the output of information normally available in short-term memory, and when only a short time elapses between the cognitive operation and the report; when the information reported is specific and pertains to a particular cognitive episode (versus a general assessment covering a long series of behaviours); and when reported information is descriptive rather than interpretive (Hamilton, Nussbaum & Snow, 1997; Touragneau, 2003; Castillo-Diaz & Padilla, 2013). An examination of the judgment-forming process would meet most if not all of these criteria. For example, think-aloud protocols occur “in the moment”, and would necessarily report the information in short-term memory being used to form judgments. Other criteria, such as the advantage of reporting on a particular cognitive episode, reflect observed judgment strategies, thus eliminating the need to impose these limitations methodologically.

The predominant criticisms of cognitive interviewing involve cases in which respondents misinterpret the reasons for carrying out a certain behavior or fail to identify correctly the stimuli that affected a particular response (Nisbett & Wilson, 1977). In the case of forming judgments, the participants’ rationale for their selection of judgment strategies is less important than their self-report of which strategies were used in the interval between the presentation of and their response to an assessment item. Errors in participants’ ability to identify the relationship between stimulus and response could provide further insights into why participants arrive at certain answers (e.g., the point at which a bias is introduced in the judgment-forming process), posing a potential benefit to the use of cognitive interviewing to understand judgment formation.
**Item Grouping**

Judgment formation is mediated by both the content of a self-assessment, and the context in which the self-assessment is completed. These factors, which influence the selection and application of judgment strategies, can influence participant’s responses in ways that affect reliability and validity of these responses. Earlier in this paper, respondent factors were examined (intentional and unintentional biases). This section will focus on the instrument and context factors that affect the validity of responses.

**Random and systematic error.** There are two types of measurement errors: random and systematic errors. Random error is the measurement error resulting from the inherent unpredictability of the interaction between the environment and the assessment process. Random errors lead to unpredictable and inconsistent responses both within a measure (such as a survey) as well as on repeated measures across time. and in both instances they and affect reliability of the obtained scores. In contrast, systematic errors generate a predictable, constant offset from the true value (i.e., consistently inflate or deflate results), and consequently affect validity of the scores.

Systematic errors in self-assessment can be controlled if their sources are identified; however, identifying and controlling for these errors can be challenging, particularly when their effects are mistaken for moderating effects. Moderating variables alter responses but such effect is not a bias effect. For example, gender can be a moderating variable for the responses to a scale designed to measure attitude towards auto mechanic in that male has a more favourable attitude than female respondents, but such a between group difference is accurate if both groups understand and interpret the items similarly. On the other hand, if the female respondents do not understand the items (because of the jargons used) as well as the male respondents, which leads
to lower reliability and consequently validity of their responses, then a bias due to gender is observed.

**Construct validity of self-assessment.** Construct validity is the degree to which an assessment measures the intended construct; and by extension, the appropriateness of inferences made about the target construct on the basis of the observed scores. In the context of self-assessment, there are two important ways in which this systematic bias can be introduced as a function of the assessment: construct irrelevant variance (CIV) involves the introduction of extraneous, irrelevant variables that affect assessment results. Examples of sources of construct irrelevant variance include poorly constructed or irrelevant items and respondents with propensity to provide socially desirable responses. Invalidity and bias can occur at the item level. Item invalidity occurs if all respondents misinterpret an item. Item bias occurs when participants in one group (e.g., female) are less likely to respond correctly to an assessment item than participants in another group (e.g., male) as a function of being a member of that group. Differential item functioning, i.e., respondents in different groups of equal ability responding in systematically different ways to the same items, result in a systematic inflating or deflating of scores of a certain group.

Another risk to the validity of self-assessments is construct underrepresentation (CU), which occurs when participants’ responses do not adequately reflect the target construct as a result of the instrument not including an array of items that measure all the different aspects of the construct (Downing, 2002). Sources of construct underrepresentation include trivial content, too few examination items, too narrow a definition of the construct, and poor distribution of assessment items (Downing, 2002).
These context-based threats to construct validity are critical in the examination of judgment formation and selection of judgment strategies, as these particular threats act on participants’ cognitive processing of the assessment items.

**Context effects.** Context effects are a type of instrument factor referring to “any influence or interpretation that a respondent might ascribe to an item solely because of its relation to the other items making up an instrument” (Wainer & Keily, 1987, p.187). These errors are a product of the context in which the items on the questionnaire are placed, and manifest themselves in a variety of ways (Perlini, Lind & Zumbo, 1998; Podsakoff et al., 2003; Weinberger, Darkes, Del Boca, Greenbaum & Goldman, 2006).

**Item-priming effects.** Item priming effects refer to an implicit memory effect in which exposure to one stimulus influences responses to other stimuli by increasing the salience of related information (Perlini, Lind & Zumbo, 1998). For example, Harrison, McLaughlin, and Coalter (1996) demonstrated that shorter scales (i.e., containing fewer items) improve the accessibility of answers to previous scales, increasing the likelihood that previous responses will influence current answers. Shorter scales are believed to minimize the decay of previous responses in short-term memory, augmenting the perceived relationships between scale items with similar content (Harrison et al., 1996). Harrison and McLaughlin (1993) observed that neutral items embedded in the context of either positively or negatively worded items take on the evaluative properties of those items, thereby affecting respondent’s interpretation of the question, information retrieval from memory, judgment of the retrieved information and ultimately, their response. These effects are believed to produce artificial covariation among variables (Salancik, 1984).
**Item grouping effects.** Building upon the concept of item-priming are the effects of grouping “like” items together. Research has established that the sequencing of items can have a significant effect on respondents’ answers, especially for more cognitively taxing assessments (Hogarth & Einhorn, 1992; Melnick, 1993; Lam, Green & Bourdignon, 2002; Ortner, 2008). This is referred to as an item order effect, in which the responses to early items on an assessment affect the response to subsequent items. Item order is a context effect believed to dictate the way in which items are processed by the respondent, creating an interactive effect between the respondent’s information processing strategies, and the task (Hogarth & Einhorn, 1992; Melnick, 1993).

Intermixing items of different constructs within the same questionnaire is common, and in some cases, a recommended practice to reduce method variance (Kline et al., 2000). If the constructs of the questionnaire are similar, however; this practice might increase the inter-construct correlations (because respondents have more difficulty distinguishing between the constructs) while reducing the intra-construct correlations (because respondents also experience more difficulty seeing the similarity in items measuring the same construct), producing an artificial covariation among the constructs and potentially reducing the reliability of the scales (Podsokof et al., 2003). It should be noted that these findings are somewhat contradictory, since a reduction in scale reliability should result in a decrease of covariation among constructs (Podsokoff et al., 2003).

Item grouping is a form of item order effect examining the effects of grouping items (e.g., by theme, construct, sub-scale) together on a questionnaire or during an interview, as opposed to randomizing all items (Kirchner & Uphoff, 1955; Melnick, 1993; Lam, Green & Bourdignon, 2002). Item grouping has received some attention in the research, albeit to a lesser extent than
item order effects. Preliminary findings have been mixed. Research in support of item grouping argues that grouping allows the respondent to focus on each area (construct) individually, rather than mentally switching between constructs while responding to questions at random (Melnick, 1993; Lam, Green & Bourdignon, 2002). The premise is that by focusing the respondent’s attention without interruption, the retrieval process is facilitated through the creation of a continuous mental set they can use as a point of reference (Melnick, 1993; Lam et al., 2002). Conversely, this continuous mental set might mislead respondent’s mental processing (including the cognitive strategies they employ) (Melnick, 1993; Lam et al., 2002). For example, it is possible that the stream of consciousness prompted by item grouping might also introduce a systematic variation that artificially increases the respondent’s estimates (Melnick, 1993; Lam, Green & Bourdignon, 2002). This variation would result in falsely elevated internal reliability because when responding to a group of items sharing the same theme, the respondents may form their judgment on the basis of the theme and not just the content of the individual item (Melnick, 1993; Lam et al., 2002). Consequently, despite increased response reliability, item grouping, item response validity is reduced. Other research has generated “no effect” results, suggesting that item grouping generally does not result in significantly different responses (Monk & Stallings, 1970).

Tourangeau and Rasinski (1988) compared the processes underlying the response to a questionnaire with the recall of information from a respondent’s memory. Generally, this process would involve activating the relevant construct based on the question, retrieving construct-relevant information, and synthesizing an answer from this retrieved content (Tourangeau, 1984). During the first phase, when respondents interpret the items presented, earlier items can provide a framework for the interpretation of later items (Sudman & Bradburn, 1983; Melnick, 1993;
Lam et al., 2002). This framework can influence the construct that is activated (e.g., attitude), shaping the respondent’s impression of the construct later items are intended to refer to (Sudman & Bradburn, 1983; Melnick, 1993; Lam et al., 2002). This item effect is exacerbated when items are grouped and the group themes are disclosed to the respondents.

Studies dealing with comprehension have demonstrated that respondents transfer the gist of their interpretation of the context (from earlier items) to memory, and this interpretation creates an imprint or framework through which the remainder of the questionnaire or sub-scale is interpreted (Bower, Black & Turner, 1979). These observations are akin to the “priming effect” known in psychology, which states that once a particular construct is activated; associated contents (e.g., attitudes, emotions) are automatically (unconsciously) activated (Anderson, 1983; Johnson & Tversky, 1983). In terms of assessment, the context set through the grouping of items is believed to affect how a judgment is made, for example, by affecting the anchors used in the scaling of individual characteristics, or in the way information is synthesized into a judgment (Melnick, 1993; Lam et al., 2002). Ambiguous items or stimuli are even more vulnerable to these context (including item-grouping) effects because respondents are more inclined to rely on the theme to which these items are grouped (Trope, 1986; Hogarth & Einhorn, 1992; Weinberger et al., 2006).

The impacts of item grouping have been detected in a number of studies that elaborate on the mechanisms through which these effects take place (e.g., Schurr & Henriksen, 1983; Melnick, 1993; Lam et al., 2002; Weinberger et al., 2006). These studies have found that in generating a response to an item within a group of items sharing the same theme, respondents are clearly influenced by the content of earlier items placed closest to the item (i.e., a proximity effect); and that these items aide respondents by clarifying the meaning of items (Knowledge &
Byers, 1996), and reducing the “don’t know” response on ambiguous items (Tourangeau & Rasinski, 1988; Lam et al., 2002).

Overall, the literature seems to agree that there are positive effects resulting from item groupings, although there remains some uncertainty as to whether the improved internal reliability generated by this methodology might be artificially inflated.

The enhanced reliability due to item grouping has consequences from a validity perspective, as discussed by Lam et al. (2002). One view predicts a decrease in validity resulting from the enhanced consistency of generalized responses, that are; however, inconsistent with the items being measured. This would be particularly hazardous in the case of ambiguous items (Todorov, 2000). The second view suggests that enhanced item response reliability will also improve the validity of responses by clarifying ambiguous items (Weinberg et al., 2006). These effects are of particular significance for attitude items, and therefore self-assessment scenarios.

Currently, the literature discussing the influences and impacts of these issues are predominantly hypothetical void of empirical evidence to demonstrate the effect of inflated reliability on validity of responses, and employing the indirect measurement of cognition (i.e., they are based on respondents giving predictable, systematic answers on two or more assessments evaluating related constructs) (e.g., Ehrling, 2007). In addition, although this literature is important and useful to demonstrate the impacts errors can have on survey estimates; they say little about what causes these errors to arise, or how to reduce these errors proactively.

These sources of error, present throughout each cognitive phase, create multiple internal and external, intentional and unintentional opportunities to affects a respondent’s processing of and response to assessment items. In other words, these factors influence respondent’s abilities to form judgments.
In sum, a variety of methodological recommendations have been made in an effort to mitigate measurement errors introduced by self-assessment, none of which do so satisfactorily. These alternatives address procedural flaws in the traditional self-assessment (e.g., Lam, 2009), errors resulting from lack of knowledge or skill on the part of the respondent (e.g., Ehrlinger, 2007), and challenges related to motivation, confidence and stakes associated with the decisions based on the self-assessment (e.g., Hayes & Dunning, 1997). None of these methods seem to address the cognitive processes that enable self-assessment (forming judgments) and their moderating factors. Although the moderating factors of many self-assessment errors have been identified, when, how, and why these factors are introduced remains unknown. An analysis of the cognitive processes involved would help to decompose the steps participants follow to form judgments, enabling a more targeted approach to reducing self-assessment errors through refinement of instrument development, and consequently improving the validity of self-assessment data.
Chapter 3: Methods

To examine the cognitive processes used to form judgments, this study employed a 1x2 group design in which doctoral students were asked to self-assess their quantitative and qualitative research skills in item-grouped (experimental) and item non-grouped (control) conditions. In both conditions think-aloud and retrospective probing techniques were leveraged during in-person cognitive interviews to uncover participants’ judgment-forming strategies. In addition to self-assessing their research skills, participants were asked to indicate whether they identified themselves as qualitative or quantitative researchers (“researcher type”), and to evaluate their overall (quantitative and qualitative together) research skills now, versus when they first began their doctoral studies.

Participants & Setting

Study participants were doctoral students at the Ontario Institute for Studies in Education (OISE), in the department of Curriculum, Teaching, and Learning (CTL). At the doctoral level, all students have completed at least one course in research methods at OISE, providing them with a common baseline understanding of the research construct (research efficacy). This baseline knowledge helped to manage individual differences in the understanding of the construct and associated measures being used in the self-assessment, and was verified through the recruitment email.

Potential participants were recruited from the pool of doctoral students in CTL via an email through the mass emailing system through the Dean’s office. The CTL department heads disseminated the research participant recruitment email through their student list-servs. Each participant was sent a copy of the Information Sheet/Consent form via email and invited to schedule a meeting time for the interview.
A sample size of 8 participants per condition was targeted (16 total subjects). Although this sample size is small in quantitative studies, previous research has validated its appropriateness for exploratory qualitative research and the use of cognitive interviewing (e.g., Krueger & Casey, 2000; Willis 2005; Patton, 2002). Willis reinforces that saturation is reached within five to nine interviews of any single type of group (2005).

Although the construct of research efficacy is being used for the assessment, the purpose of this study is not to examine this construct; instead, it is being used exclusively as a medium through which to examine participants’ cognitive processing. The demographic makeup of the participants, therefore, beyond being doctoral students is less meaningful. For this reason, securing a random or representative sample of the broader population was not a significant priority and logistically impossible within the context of this research, although background information was collected to explore potential moderating variables. Pre-existing knowledge is a known moderating variable, controlled through the exclusive inclusion of doctoral students.

Table 1 provides a summary of participant demographics.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Condition</th>
<th>Researcher Type</th>
<th>Year of Study</th>
<th>Dissertation Defended?</th>
<th>Research Studies (investigator)</th>
<th>Research Studies (Subject)</th>
<th>Research Courses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Non-Grouped</td>
<td>Quantitative</td>
<td>2</td>
<td>Y</td>
<td>10+</td>
<td>10-15</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Non-Grouped</td>
<td>Qualitative</td>
<td>4</td>
<td>N</td>
<td>3</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>Item-Grouped</td>
<td>Qualitative</td>
<td>6</td>
<td>N</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Item-Grouped</td>
<td>Qualitative</td>
<td>1</td>
<td>N</td>
<td>4</td>
<td>5-10</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Non-Grouped</td>
<td>Qualitative</td>
<td>4</td>
<td>N</td>
<td>3</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>Non-Grouped</td>
<td>Qualitative</td>
<td>4</td>
<td>N</td>
<td>1-3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>Non-Grouped</td>
<td>Qualitative</td>
<td>5</td>
<td>Scheduled</td>
<td>3-5</td>
<td>5-10</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>Item-Grouped</td>
<td>Qualitative</td>
<td>4</td>
<td>Scheduled</td>
<td>4</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>Item-Grouped</td>
<td>Qualitative</td>
<td>5</td>
<td>Scheduled</td>
<td>3-5</td>
<td>5-10</td>
<td>3</td>
</tr>
<tr>
<td>Item</td>
<td>Grouped</td>
<td>Qualitative</td>
<td>N</td>
<td>Range</td>
<td>Range</td>
<td>Range</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>-------------</td>
<td>---</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Item</td>
<td>Qualitative</td>
<td>5</td>
<td>N</td>
<td>3-5</td>
<td>5-10</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>Item</td>
<td>Qualitative</td>
<td>3</td>
<td>N</td>
<td>2</td>
<td>3-5</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>Non</td>
<td>Quantitative</td>
<td>2</td>
<td>N</td>
<td>2</td>
<td>3-5</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>Non</td>
<td>Qualitative</td>
<td>6</td>
<td>N</td>
<td>4</td>
<td>5-10</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>Item</td>
<td>Qualitative</td>
<td>4</td>
<td>N</td>
<td>3-5</td>
<td>5-10</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>Item</td>
<td>Qualitative</td>
<td>1</td>
<td>N</td>
<td>3-5</td>
<td>3-5</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>Non</td>
<td>Qualitative</td>
<td>1</td>
<td>N</td>
<td>2</td>
<td>3-5</td>
<td>2</td>
</tr>
</tbody>
</table>

**Year of Study.** All sixteen participants were doctoral students in the department of Curriculum, Teaching, and Learning at OISE. Of those sixteen, three were in their fifth year (18.6%) and had defended or would be defending their dissertation within the month; two were in the later stages of their degree (beyond year 5) (12.5%), but would not be defending in the near-term; six were mid-way through their degree and had completed all of their required courses (37.5%); and four were in the first two years of their doctorate (25%), in the process of completing their course requirements. One participant had completed all courses and defended their dissertation in two years (6.25%).

**Research Experience – as Researchers.** This study focused exclusively on doctoral candidates, ensuring all participants had conducted primary research for at least one research project. The majority of participants (62.5%) had acted as senior researchers for 3-5 research projects. Four participants (25%) were researchers for 2 research projects. One participant indicated that their doctoral research was the only research study for which they had been a researcher; and one final participant reported being involved in ten or more research studies as a researcher. Research experience for all but one participant was limited to qualitative research studies.

**Research Experience – as Participants.** Participation in peer research studies is
common for graduate students. Of the sixteen participants, only two reported this study as the first study they had participated in (12.5%). Two participants indicated they had participated in ten to fifteen studies as a participant (12.5%). The majority of participants (75%) reported involvement in approximately five research studies as a subject.

**Research Methods Courses.** Completion of at least one research methods course is a mandatory requirement for completion of a doctoral degree at OISE. As such, all participants had completed at least one research methods course through CTL. Eight participants (50%) had completed a total of three methods courses, all at OISE; while another four (25%) had completed two research methods courses at OISE. Two participants had taken only one research methods course (12.5%), and did so at OISE. Only two participants took research methods courses outside of OISE; both took two courses outside of OISE, and two or three research methods courses at OISE. All participants had taken a mixed methods course at OISE. Participants who took two or more research methods courses had taken qualitative courses. Two participants had taken one quantitative research methods course each.

**Defining “Research Experience”.** For the purposes of this analysis and the subsequent discussion, “experience” was defined first by participant’s direct research experience (i.e., the number of research studies they had participated in as researchers), and then by progress in their degree (i.e., course and dissertation completion). Participants were categorized as having “high”, “moderate”, or “low” research experience. This differs slightly from the original thinking, which focused on participant’s year of study as commensurate with research experience. Through the interviews, it was revealed that in fact some of the most “experienced” participants had been in the program for less than five years, but had been involved in more research as an investigator than participants in the program for more than five years.
Materials

The self-assessment items were developed based on the Clinical Research Appraisal inventory (Mills, Caetano & Rhea, 2014). The original Clinical Research Appraisal inventory includes 92 items representing 10 thematic areas. Two areas (qualitative and quantitative research skills) were used for this study, totalling 10 self-assessment items. Although this inventory was initially developed for use with physicians, it has been tested with undergraduate and graduate students, and several of these categories were deemed not applicable for this population (Eller & Bakken, 2014). An additional two self-assessment items were included, which asked participants to rate their level of confidence in their research skills now, and before beginning their doctoral studies at OISE.

The self-assessment forms were divided into two parts. The first section included background questions inquired about participant’s gender, year of study at the doctoral level, the number and type of research methods courses taken at OISE, those taken at other institutions, and the number of research studies they had participated in as researchers and as subjects. The second section included the question items. In the non-grouped (control) condition, this section included a randomized set of research skills items about qualitative and quantitative research skills. In the item-grouped (experimental) condition, items were divided into quantitative and qualitative sub-sections. Each sub-section began with the category definition (quantitative research skills and qualitative research skills), and included 5 questions. The self-assessment instruments may be reviewed in Appendix D.

Design

This study employed a 2x2 mixed factorial design with one between-subjects factor: item grouping (item-grouped vs. non-grouped) and one within-subjects factor: item type (quantitative
vs qualitative). The item non-grouped or control group, consisting of 8 doctoral students, was presented with the survey in which qualitative and quantitative items were mixed. The item grouped or experimental group, consisting of 8 doctoral students, was presented with a survey item-grouped in which the qualitative and quantitative assessment items were separated. Participants in this item-grouped condition were asked to evaluate their quantitative research skills first, and their qualitative research skills second. In both conditions think-aloud and retrospective probing strategies were leveraged during in-person cognitive interviews to uncover participants’ judgment-forming strategies.

The following measures were used in both conditions: mean self-assessment scores, mean qualitative scores, mean quantitative scores. Additionally, gender, researcher type, year of study, number of research studies (as a participant), number of research studies (as a researcher), and overall confidence in research skills were also evaluated.

**Data Collection & Analysis**

Interviews, including debriefs, were recorded. Hand-written notes by the interviewer were also taken during the interviews. Together, these served as the data for this study.

Data analysis sought to identify patterns across interviews related to the cognitive steps followed by participants to formulate judgments. The data was also analyzed in the context of item grouping (e.g., participants might make reference to the grouped items to evaluate their research skills (item-grouped scenario – Form B)), in order to understand whether and how the effects of item grouping influenced judgment strategies and outcomes.

The data was coded according to the eight judgment strategies identified by Lam (2014). Coding criteria are presented in Table 2. Multiple judgment strategies were coded as “major” and “secondary” strategies. Participants’ “major” strategy was noted as that which they use most
elaborately and which generates their assessment score. “Secondary” strategies were identified when participants leveraged additional strategies following the major strategy. Secondary strategies were not used as elaborately (i.e., participants were not expected to spend as much time discussing them), and did not significantly alter the participant’s assessment score.

Table 2. Major Judgment Strategies

<table>
<thead>
<tr>
<th>Judgment Strategies</th>
<th>Associated Concepts/Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific Event</td>
<td>Use of only 1-2 specific events as point(s) of reference for judgment formation.</td>
</tr>
<tr>
<td>General Pattern</td>
<td>Reference to a pattern of thoughts, behaviours, or outcomes (i.e., “I tend to be good at writing tests”).</td>
</tr>
<tr>
<td>Other’s Opinion</td>
<td>Explicit reference to another person’s opinion as a proxy for self-evaluation (e.g., “my professor thinks/says…”).</td>
</tr>
<tr>
<td>General Trait</td>
<td>Reference to a general trait possessed by the participant (e.g., “I am a good student”)</td>
</tr>
<tr>
<td>Consequences</td>
<td>Reference to consequences (e.g., “I’m successful”) and inductive references to meaning (e.g., “I’m successful so it must mean that…”).</td>
</tr>
<tr>
<td>Feeling</td>
<td>Reference to feelings related to or experienced when performing the target behaviour (e.g., “I enjoy…”). References are intuitive and emotion-based, versus objective.</td>
</tr>
<tr>
<td>Analysis</td>
<td>Breakdown of target behaviour into component parts and generation of specific examples that examine each part. Examples are diverse (i.e., drawn from multiple experiences). Participants make a concerted effort to examine both supporting and disconfirming evidence of the target behaviour.</td>
</tr>
<tr>
<td>Socially Desirable Response</td>
<td>Reference to participants’ identity, explicit reference.</td>
</tr>
</tbody>
</table>

Procedure

A pilot test group was conducted with three subjects from the pool of volunteers. Results from this pilot were used to test item comprehension, which resulted in no revisions to the procedure and questions described herein.

Interviews began by revisiting the informed consent provided by participants to ensure their understanding and to provide them with an opportunity to withdraw from the study if they desired. Upon confirmation to proceed, each interview included two phases: during phase one, participants were asked to answer a set of self-assessment questions on research efficacy using the “think aloud” method (concurrent verbal protocol). This approach required participants to
verbalize the inner language activity they experience while responding to the self-assessment.

During this phase, a series of follow-up probes, such as “What were you thinking about when…” were used to more thoroughly explore participants’ approaches to formulating an assessment response see Appendix B for the interview script and sample questions). Specifically, these follow-up probes allowed the investigator to inquire about aspects of the think-aloud that the participant might not have verbalized or pursued in detail.

Participants were provided with a final debrief at the end of phase two (see Appendix C). This debrief explicated the study’s design and deception regarding the item grouping. At this time, participants were given the option to withdraw from the study and, if so, advised that their data would be discarded. No participants chose to withdraw from the study at any time.
Chapter 4: Results

The results of this study are organized by first discussing background variables, to set context for the analysis. Next, each of the cognitive strategies participants used to respond to the assessment items are described. Distinctions are made between main cognitive strategies (strategies used to form initial impressions and provide preliminary assessment responses for each question), and secondary strategies (strategies employed when participants elaborated on their assessment responses in response to follow up questions). Participants’ self-identification as qualitative or quantitative researchers is explored as the end to provide context for a discussion regarding item grouping. Finally, ratings on the items pertaining to quantitative and qualitative research skills for the item-grouped and item-non-grouped conditions were analyzed using the Two-Way mixed factorial design.

Direct interview quotes are used to add depth to the data. Quotes have been edited for grammatical clarity and all names and identifying characteristics (e.g., specifics about participant’s research projects) have been changed to protect their identity.

Mean Research Skills Scores and Research Confidence

The mean self-assessment score for qualitative items was high (8.5 out of 10), while the mean score for quantitative items was moderate (6 out of 10). These scores remained consistent regardless of the treatment condition. This is in-line with previous research on graduate student research skills (e.g., Meerah, Osman, Zakaria, Ikhsan, Krish, Lian & Mahmod, 2012). Table 3 shows descriptive statistics of ratings of all items, and quantitative and qualitative items for the item-grouped and item-non-grouped conditions.
Table 3. Summary statistics by question for non-grouped and item-grouped conditions.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Std. Error</th>
<th>95% Confidence Interval for Mean</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>16</td>
<td>7.08</td>
<td>1.39</td>
<td>.35</td>
<td>6.34</td>
<td>7.82</td>
<td>4.80</td>
</tr>
<tr>
<td>Self-Assessment</td>
<td>8</td>
<td>8.00</td>
<td>.99</td>
<td>.35</td>
<td>7.17</td>
<td>8.83</td>
<td>5.60</td>
</tr>
<tr>
<td>Item-Grouped</td>
<td>8</td>
<td>6.16</td>
<td>1.09</td>
<td>.39</td>
<td>5.25</td>
<td>7.08</td>
<td>4.80</td>
</tr>
<tr>
<td>Qualitative Score</td>
<td>8</td>
<td>8.68</td>
<td>.49</td>
<td>.17</td>
<td>8.27</td>
<td>9.08</td>
<td>7.80</td>
</tr>
<tr>
<td>Non-Grouped</td>
<td>8</td>
<td>8.60</td>
<td>.48</td>
<td>.17</td>
<td>8.20</td>
<td>9.00</td>
<td>8.00</td>
</tr>
<tr>
<td>Item-Grouped</td>
<td>8</td>
<td>8.60</td>
<td>.48</td>
<td>.17</td>
<td>8.20</td>
<td>9.00</td>
<td>8.00</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>8.64</td>
<td>.47</td>
<td>.12</td>
<td>8.39</td>
<td>8.89</td>
<td>7.80</td>
</tr>
<tr>
<td>Quantitative Score</td>
<td>8</td>
<td>7.33</td>
<td>2.03</td>
<td>.72</td>
<td>5.63</td>
<td>9.02</td>
<td>2.60</td>
</tr>
<tr>
<td>Non-Grouped</td>
<td>8</td>
<td>7.33</td>
<td>2.03</td>
<td>.72</td>
<td>5.63</td>
<td>9.02</td>
<td>2.60</td>
</tr>
<tr>
<td>Item-Grouped</td>
<td>8</td>
<td>3.73</td>
<td>2.50</td>
<td>.88</td>
<td>1.63</td>
<td>5.82</td>
<td>1.00</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>5.53</td>
<td>2.88</td>
<td>.72</td>
<td>3.99</td>
<td>7.06</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Two patterns of responses were revealed during the interviews: 75% of participants provided scores ranging from mid- to high-proficiency (6-9 out of 10) for all items, regardless of their personal identification as a qualitative or quantitative interviewer or the treatment condition. This is interesting given that participants who self-identified as qualitative interviewers frequently made reference to their lack of knowledge regarding quantitative skills and procedures and yet their ratings were higher than the mid-point of the rating scale. The second pattern was a polarizing of scores – participants reported high scores on the items that fit their self-identification as either qualitative or quantitative, and reported very low scores for items that were outside of that identification (i.e., qualitative researchers rated themselves very high on qualitative items and very low on quantitative items). While this polarization was observed across treatment conditions, because in most cases the participant was able to identify whether a skill was more qualitative or quantitative, the extremity of the scores were more pronounced in the item-grouped condition.
All participants rated their overall research efficacy as very high (between 8-9 out of 10), and their efficacy prior to OISE as high (7-8 out of 10). Interestingly, participants noted a difference between the confidence levels they experienced at the commencement of their degree, and whether that level of confidence was justified based on their current (expanded) understanding. In other words, many participants rated their pre-OISE confidence as high (7 or 8 out of 10), but in retrospect (after self-assessing their research efficacy) can acknowledge that this confidence was misguided because they now know much more and would rate themselves around the same.

87.5% of participants responded quickly to the overall research efficacy question and were unable to articulate their strategy for arriving at their self-reported score. With further probing, participants tended to reference the same strategies as were used for the skills-based assessment questions. It is suspected that the act of completing the self-assessment just prior to evaluating their overall confidence in their research skills provided an intuitive sense for their “average” capability level, reflecting an item-order priming effect. Two participants made direct reference to taking the approach of finding the mean, stating:

“Well, I looked at the scores I’d given myself on all of the items, and the average looks to be at about an eight. So I’d say I’m at an eight in my research confidence. When I first started here, my confidence was probably about the same, but maybe it wasn’t right to feel that way. I know more now than I did then, but I feel like I would have been capable of figuring it out back then as well. So maybe my confidence hasn’t really changed during the program but my proficiency has.”

**Judgment Strategies**

The analysis of participants’ narrated self-assessment data revealed the eight distinct judgment strategies as presented in the earlier section of this report that are reflected in the current literature (Lam, 2014): (1) refer to a specific event (specific event), (2) to a general
pattern regarding the target attribute (general pattern) (3) refer to other’s perception (other’s opinion), (4) refer to a related general trait (general trait), (5) refer to a consequence of the attribute (consequences), or (6) to an associated feeling (feeling), (7) break an attribute down into specific elements and evaluating each element (analysis), and (8) rely on a socially desirable response (socially desirable response). Table 4 summarizes the main judgment strategies employed by participants.

Table 4. Summary of main and secondary judgment strategies employed by participants

<table>
<thead>
<tr>
<th>Condition</th>
<th>Researcher Type</th>
<th>Research Experience</th>
<th>Mean Self-Assessment Score</th>
<th>Mean Qualitative Score</th>
<th>Mean Quantitative Score</th>
<th>Main strategy</th>
<th>Secondary strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item-Grouped</td>
<td>Qualitative</td>
<td>Low</td>
<td>5.9</td>
<td>9.2</td>
<td>2.6</td>
<td>Specific event</td>
<td>General trait Socially desirable</td>
</tr>
<tr>
<td>Item-Grouped</td>
<td>Qualitative</td>
<td>High</td>
<td>7.9</td>
<td>8</td>
<td>7.8</td>
<td>Analysis</td>
<td>General pattern Specific event</td>
</tr>
<tr>
<td>Item-Grouped</td>
<td>Qualitative</td>
<td>Moderate</td>
<td>5.9</td>
<td>9.2</td>
<td>2.6</td>
<td>General pattern</td>
<td>Analysis Feeling Other’s opinions</td>
</tr>
<tr>
<td>Item-Grouped</td>
<td>Qualitative</td>
<td>High</td>
<td>5.4</td>
<td>9</td>
<td>1.8</td>
<td>Specific event</td>
<td>General pattern</td>
</tr>
<tr>
<td>Item-Grouped</td>
<td>Qualitative</td>
<td>High</td>
<td>6.2</td>
<td>8.2</td>
<td>4.2</td>
<td>Analysis</td>
<td>Specific event General pattern</td>
</tr>
<tr>
<td>Item-Grouped</td>
<td>Qualitative</td>
<td>Moderate</td>
<td>4.8</td>
<td>8.6</td>
<td>1</td>
<td>Socially desirable</td>
<td>Consequences</td>
</tr>
<tr>
<td>Item-Grouped</td>
<td>Qualitative</td>
<td>High</td>
<td>7.7</td>
<td>8.2</td>
<td>7.2</td>
<td>Specific event</td>
<td>Feeling Other’s opinions</td>
</tr>
<tr>
<td>Item-Grouped</td>
<td>Qualitative</td>
<td>Moderate</td>
<td>5.5</td>
<td>8.4</td>
<td>2.6</td>
<td>Associated feeling</td>
<td>General pattern Socially desirable</td>
</tr>
<tr>
<td>Non-Grouped</td>
<td>Quantitative</td>
<td>High</td>
<td>8.4</td>
<td>8.4</td>
<td>8.4</td>
<td>Analysis</td>
<td>General pattern Specific events</td>
</tr>
<tr>
<td>Non-Grouped</td>
<td>Qualitative</td>
<td>Moderate</td>
<td>8.1</td>
<td>8.8</td>
<td>7.4</td>
<td>Specific event</td>
<td>Analysis General pattern Other’s opinions</td>
</tr>
</tbody>
</table>
Further, all participants demonstrated a tendency to leverage secondary strategies to expand upon their initial judgment. It is worth noting that in all cases, the secondary strategy was used to confirm participants’ self-assessment by drawing upon other supporting examples. None of the participants’ self-assessments were altered by their use of secondary strategies (i.e., information revealed through the application of a secondary judgment strategy did not change the participants’ self-assessment score). When participants’ self-assessment scores were modified, it was always through the continued application of their main judgment strategy.

**Referring to a Specific Event.** 68.8% of participants referred to a specific event, usually their dissertation research, during the think-aloud. Of those eleven, 63% used this strategy as their major strategy (i.e., the strategy applied to generate an initial response). Some participants were explicit in citing their dissertation research as the point of reference,
“My dissertation involved working directly with my participants and being part of their day-to-day interactions…it’s very unique research…it was participatory…so I mean, obviously I know how to observe. And I take copies notes. I have a thousand pages of notes all hand-written because it’s more natural than sitting there with a computer and having my participants wonder what I’m doing.” (self-rating: 9/10)

While for others, the fact that they were referencing a single project became clear through their narrative.

25% of participants used this approach to expand upon or rationalize their answers after providing an initial response. One participant drew upon their dissertation research experience after deciding they excelled at coding,

“Coding comes very naturally to me. It feels easy and I feel good doing it. The research I’m doing right now, that’s how coding feels. I get into a zone and I can just sit there and do it all.” (self-rating: 8/10)

Participants who employed this strategy were all in the middle stages of their doctoral program (had completed course work, comprehensive exams, and were conducting research). This strategy might be a reflection of the way these participants tended to be spending their time, creating a recency effect that guided their choice of strategy. Alternatively, in some cases, this was the only research they had conducted as investigators, and was thus the only point of reference they were able to use.

**Referring to a general pattern regarding the target attribute.** 68.8% participants referred to a general pattern of behaviour(s) related to research skills as part of their evaluation strategy. However, only 12.5% of participants used this strategy as their primary judgment strategy. 56.3% of participants used this as a secondary strategy, after providing an initial assessment.

“I selected the population I would be targeting in my research and I think I did a pretty good job of it…when I think about other people’s research or I read studies, I’m not
surprised by who they included in the study so I think overall I’m good at it.” (self-rating: 8/10)

“I did live observation so I had to be really good with my notes. I’m actually just really good at taking notes in general. People tell me that all of the time. I take the best minutes for the student government I’m on.” (self-rating: 10/10)

The direct relevance of the general pattern being referenced varied by participant. While some participants seemed to be drawing on research-specific examples to support their thinking, others drew parallels to similar skills used in different contexts. For those participants in the latter group, there seemed to be no acknowledgement that the skills might be easier or more challenging to use based on the context (i.e., they might not be equally good under all circumstances). Surprisingly, the research experience of participants leveraging this strategy varied, suggesting that more experience did not necessarily support a more specific approach to assessment.

**Referring to other’s perceptions.** 43.8% of participants made reference to other’s opinions and perspectives of their research skills during their self-assessment. In all seven of these cases, this strategy was used secondarily, and was often used by participants who also anchored their assessment on a general pattern.

“I had to set up and test hypotheses because I’ve done a lot of quantitative research. In fact all of my research has been quantitative. This is something I’m good at. I’ve had a lot of feedback telling me I’m good at this, and I’ve had peers ask me for help with this.” (self-rating: 9/10)

“I’m working on a research project right now for my advisor and we are recording the interviews. She wanted me to do the coding because she feels I’m good at it, so I guess that means that I’m strong in coding.” (self-rating: 8/10)

**Referring to a related general trait.** Participants who refer to a general trait to judge their research skills tend to reflect on a broad trait, such as organization, strong academic
performance, problem-solving skills, to inform their assessment of the quality of their performance on a specific research skill or task. 12.5% of participants used this approach to rationalize scores of 6/10 for a skill they had never used:

“Well I haven’t ever used quantitative software so I don’t really know…but I’m really good at figuring things out and I’m a fast learner so I’d probably figure out how to use it and it wouldn’t take me long. So I’d put myself at a six.” (self-rating: 6/10)

As with references to general patterns, the traits referenced by participants varied in their alignment and specific relevance to the research skills being explored. Additionally, research experience and program year varied across participants using this strategy, suggesting that something other than access to relevant knowledge is informing the strategy selection of these participants.

**Referring to a consequence of the attribute.** 18.8% of participants made reference to the outcomes or consequences of their research skills as rationale for their self-assessment. Interestingly, the consequences referenced were not always direct outputs of the skill being evaluated. One participant referenced the grade she received in a mixed-methods course to rationalize their self-assessment of a skill they had not applied to an actual research project,

“I haven’t done any research with actual hypotheses so I don’t have to identify all the different variables the way you do in quantitative studies. I took the mixed methods course here [at OISE] though, and I got an A+ so I think I’d be able to do it. I’d say I’m about a seven.” (self-rating: 7/10)

In other cases, the examples used were directly relevant, and the participant used the positive outcomes they achieved as a proxy for an evaluation of the quality of their performance. One participant used the feedback and insights received from participants to assess the quality of their interview questions,

“I’m not sure I think I’m probably an eight at interview questions because I had to use
them in my research and actually you know what I got really good insights. My participants had a lot to say and it was the information I was looking for so I think that means I did a good job.” (self-rating: 8/10)

Participants who chose indirect outcomes to evaluate the skill tended to have less direct research experience, and to be in the earlier stages of the doctoral program. Participants, who chose appropriate outcomes, in contrast tended to have more research experience and/or were in the latter stages of their program. It appears that an understanding of quality, not only an understanding of the behaviours involved in a specific research skill, is required to leverage this judgment strategy appropriately. This knowledge constraint is not unlike the limitations observed in taking an analytical approach to forming judgments (discussed next).

Breaking the attribute down into specific elements and evaluating each element.

50% of participants (n=8) took an analytical approach to their self-assessment, either as a primary (n=4) or secondary (n=4) strategy. This strategy involves breaking a skill down into its component parts and assessing each of those parts. Participants then tended to report what they perceived to be the average of those smaller scores to evaluate the overall skill or attribute. One participant took a very meticulous approach to evaluating their statistical analysis skills,

“Well when you do statistical analysis you need to have to know stats – like regression, t-tests, chi-test, all of those things. I did a lot of stats in my undergrad and Master’s so I’m comfortable with that. Because yeah, the software does the calculations but you still have to actually know which ones you need and importantly what the numbers mean that it spits out. If you don’t know how to interpret those then you aren’t really any further along in your analysis. So my stats understanding is strong. Then you have to think about the statistical software. I’m assuming we are talking about quantitative software and I’ve used a bunch of those. I’ve done qualitative software too but I’m most comfortable with quantitative and I think that’s what you’re talking about. So you know, SPSS, “R”, SAS. I like using R. It’s just like SPSS but it’s better. I used it for my research here and for my Master’s and I did a bunch of other projects at a research lab so I have a lot of experience with this. I feel confident in my skills.” (self-rating: 9/10)

For this strategy, it became clear throughout the interviews that a strong understanding of
the skills being evaluated is required in order to perform an analysis of its component parts. This makes sense, given that without an in-depth knowledge of the skill being assessed (for example, participants who staunchly consider themselves qualitative researchers having to report on statistical analysis skills) participants would lack the context and information to perform such an analysis.

The participants who leveraged analysis as their primary judgment strategy were among the most experienced in research as both participants and researchers; and also tended to have greater exposure to research methods courses. Participants’ year of study was less relevant to experience, since being more advanced in one’s doctoral studies does not guarantee more hands-on research experience. This was true for this study. The most experienced participant had only been in the doctoral program at OISE for two years, and was defending their dissertation within the month; while the participant who was in their sixth year of the doctoral program had only been involved in one research project (ever) and had taken one research methods course.

Consequently, their exposure to research and practical application of their research skills was more limited than the most experienced participant, in spite of the time spent in the program.

**Referring to an associated feeling.** 31.3% of participants referred to an associated feeling when performing their self-assessment. 40% of those participants used this as a primary strategy in response to assessment questions, while the remaining participants leveraged this strategy subsequent to forming an initial evaluation. This approach to self-assessment involves reflecting on a feeling the participant seems to experience during or about this skill, and using this feeling to evaluate their confidence. One participant who used this as their primary judgment strategy used the ease with which tasks were performed as a proxy for their skill level.

“I really enjoy observing and interviewing participants. It’s fun! I like talking to people and learning about them so this is natural for me. It’s easy. I’m good at the interview
Another participant used their discomfort about using this skill to guide their assessment when they were unable to break the skill down into its component parts for analysis.

“Oh boy, that’s a quantitative question and I’m really not a numbers person. So what do you mean statistical software? I don’t even know really what you put into that. I know it’s about correlations? What else? I don’t do math (laughter)…even thinking about it, I’m uneasy. So I don’t think I’d be very good at it. I mean, I could probably learn, but it would be really hard I think so I wouldn’t feel confident even if I was doing it right.”

As with the earlier strategies discussed, variation in the research experience of participants using this strategy was observed. While in some cases, it appeared this strategy served as a proxy for lack of understanding or knowledge (of the target skill/attribute); in others it seemed to be a function of the participant using their “intuition”. It is possible this intuition is based on sound evidence supporting the participant’s self-assessment, and they were simply unable to clarify their thoughts enough to report them (for example, perhaps they occurred too quickly).

**Relying on a socially desirable response.** The final strategy used by participants involved deferring to a socially desirable response; or a response that would enable the participant to keep their self-perception as a good researcher in tact. Typically, this behaviour was observed for questions that the participant felt they would score low on; often, this strategy was revealed after a pause in the participants’ stream of thoughts. This pause often followed an attempt to break down a skill or as after drawing upon a specific example. Participants using this strategy tended to shift from reporting their current or actual confidence in a research skill, to a hypothetical confidence – what they would be if they possessed this skill. Interestingly, the absence of this skill was not reflected in their self-assessment; that is, participants consistently reported a score over 6 for these skills in spite of acknowledging that they had not used them.
One participant became very flustered upon recognition that they had not used the skill they were being asked to evaluate.

Participant: “Oh you mean doing statistics? Qualitative (Quantitative?) statistics? No. No, I don’t do that…”

Researcher: “What are you thinking about?”

Participant: “Well just that I could do it. I’m a good researcher, I’m smart. I’m sure I would be good at it. 8. I would be an 8.”

As can be observed in the results, none of the sixteen participants leveraged only a single strategy. Typically, a main strategy would be leveraged to form an initial impression in response to an assessment item, and would then use another strategy or strategies to rationalize their assessment.

**Researcher type.** 87.5% of participants (n=14) indicated they considered themselves qualitative researchers (Table 4). Many went on to describe their rationale for this self-identification, which typically focused on outlining the various research projects they had participated in and the types of activities they completed as part of these projects. 12.5% of participants identified themselves as quantitative researchers, and employed a similar strategy to rationalize their choice. Interestingly, all participants took an analytical approach to answer this question (i.e., they broke down the attribute / category and explored supporting evidence). This seems to support the possibility that improved knowledge and understanding encourages a more analytical approach to assessment. That is, when depth of knowledge is readily available, it is more likely to be used to inform assessment; when it is unavailable, alternative strategies are necessarily employed.

Overall, this question seemed important to all participants, and most were able to generate a response almost instantly. It is possible that the increased motivation also informed the use of an analytical judgment strategy, as participants were more motivated to spend the extra
time required to conduct the analysis using the information they had on-hand.

**Research experience.** Research experience was the only other factor that seemed to influence participants’ choice of judgment strategy (Table 4). All participants with “high” research experience (i.e., being involved in more than 10 research projects as an investigator and/or participant) took an analytical approach to their self-assessment. In each case, participants were able to generate a variety of diverse examples (i.e., not only their dissertation research) from a number of different contexts (e.g., working in a lab, being an undergraduate student) and use these examples to examine each component of the deconstructed attribute. The total number of participants who leveraged analysis as their judgment strategy and possessed a high level of research experience was too small to demonstrate significance in this study; however, research experience reasonably explains the depth of knowledge that facilitated these participants’ systematic and thorough approach to self-assessment.

**Self-assessment score.** Choice of judgment strategy did not vary significantly by self-assessment score. In other words, research efficacy was not a significant factor in determining participants’ approach to forming judgments about their research efficacy. This suggests that being strong in the attributes being assessed does not translate into self-assessment accuracy, or reliability in their choice of judgment strategies. This is an interesting finding in light of the previously discussed results suggesting that breadth and depth of research experience provide a better context in which to form judgments of research efficacy.

**Item-Grouping**

Two conditions were used in this study: Non-grouped condition included a randomized item list of quantitative and qualitative research skills, and the item-grouped condition leveraged a item-grouped list, in which each category was defined (i.e., qualitative versus quantitative) and
all subsequent questions pertained to that category of skills. Table 5 and Figure 1 summarize the mean scores for each item type (qualitative and quantitative) by treatment condition.

**Table 5. Summary of Qualitative and Quantitative Means by Treatment Condition**

<table>
<thead>
<tr>
<th>Item Type</th>
<th>Treatment Condition</th>
<th>Qualitative</th>
<th>Quantitative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-grouped</td>
<td>8.68</td>
<td>7.33</td>
<td>8.00</td>
</tr>
<tr>
<td></td>
<td>Grouped</td>
<td>8.60</td>
<td>3.73</td>
<td>6.16</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>8.64</td>
<td>5.53</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Mean Qualitative, Quantitative, and Total Scores by Treatment Condition

Results from the 2X2 mixed ANOVA reveal significant main effect for both factor and significant interaction effect. There was a significant main effect of item type, F (1, 14) = 23.97, p<.001, partial Eta Square = .631. This suggests that regardless of condition, quantitative items were rated significantly higher than qualitative items. A significant main effect of condition was also found, F (1,14) = 7.686, p<.05, partial Eta Square = .354, indicating that, in addition to the main effect of item type, condition had a significant effect on the way each item type was rated. Specifically, mean ratings from Non-Grouped condition were significantly higher than that from
the Item-Grouped condition. Predictably, there was a significant interaction between the item type and the condition, $F(1,14) = 12.343$, $p<.05$, partial $\eta^2 = .469$. Specifically, ratings were similar for qualitative items in both conditions, while quantitative items were rated significantly lower in the grouped condition.

The item-grouped condition seemed to invoke participant’s identification as either qualitative or quantitative researchers earlier on in the interview process; identifying as qualitative or quantitative was the last question asked in both conditions, but in the item-grouped condition, this identification was made voluntarily by the participant within the first 2-3 questions.

Following this identification, polarized self-assessment scores were observed more frequently, with direct references made to their self-identified researcher type. Moreover, participants in this condition applied greater effort to assess questions they felt were relevant to them (i.e., matched the type of researcher they were), and often did not articulate any judgment strategy for question items that did not match their research category. In other words, participants were more likely to default to a general trait (being qualitative or quantitative) to generate scores for questions in the other category. These differences in self-assessment scores can be observed in Table 5, which illustrates that the item-grouped condition generated significantly different results on the four quantitative questions (Q6: I am able to set up and test hypotheses using numerical data. Q7: I am able to select a participant sample that is representative of a population by using established sampling procedures, Q8: I can use software to conduct statistical analysis, Q9: I am able to identify dependent, independent, and extraneous variables in designing a research study), as compared to the randomized condition. Given that the majority of participants self-identified as qualitative researchers, this difference suggests an interactive effect between
researcher-type and question-type. Qualitative researchers rated themselves significantly lower on quantitative items only in the item-grouped condition.

Interestingly, research experience seemed to override item grouping in terms of judgment strategies used. Regardless of condition, participants with more research experience tended to take an analytical and detailed approach to self-assessment (see Table 4). This was true despite participants’ ability to readily identify which questions were qualitative versus quantitative, and regardless of their self-categorization.
Chapter 5: Discussion

This study identified and examined eight distinct cognitive strategies employed by participants to complete a self-assessment of their research efficacy and their apparent relationship to personal variables (e.g., research experience, year of study, designation as a qualitative or quantitative researcher) and instrument variables (i.e., grouped or randomized items). These strategies, originally identified by Lam (2014), included: referring to a specific event, to a general pattern regarding the target attribute, to other’s perception, to a related general trait, to a consequence of the attribute, or to an associated feeling, breaking an attribute down into specific elements and evaluating each element, and relying on a socially desirable response.

While the judgment strategies presented here are seldom explored in self-assessment literature, a breadth of relevant research exists in the area of decision-making that examines participant’s cognitive strategies for arriving at a decision. Given the alignment between decision-making and forming judgments, this research presents possible cognitive explanations for the biases and cognitive heuristics observed in self-assessment research, which lead to systematic reductions in assessment reliability and validity.

The current study presents a unique opportunity to explore the cognitive processes discussed in the decision-making literature from a qualitative perspective. To-date, this qualitative perspective, which examines the experience of leveraging judgment strategies (versus examining only the outcomes of these choices), is largely absent from the literature. This discussion will therefore serve as a preliminary attempt to infuse a qualitative perspective to expand upon the current research in self-assessment, cognition, and decision-making.

Q1: What cognitive strategies do respondents use to evaluate the specific aspects of their research efficacy?
Research in the areas of decision-making and self-assessment reflect two orienting perspectives: one perspective examines the underlying processes involved in decision-making and/or self-assessment, while the other considers the knowledge and/or skills required to move through these processes, and the extent to which participants’ knowledge and/or skills affects the efficacy and accuracy of these mental processes. The results of the current study may be interpreted from both perspectives, and I will begin with a process-focused analysis.

**Dual Process Theories**

The results of this study align with research on dual process theories of reasoning (Epstein 1994; Evans 1984; 1996; Evans & Over 1996; Sloman 1996). The eight judgment strategies observed during the cognitive interviews of this study may be argued to reflect both systems of processing, and can be categorized accordingly. These differences are perhaps best described, quite literally, by Kahneman’s (2001) references to fast (System 1) and slow (System 2) thinking. For example, participants who referred to a feeling associated with the target attribute generated rapid, cursory scores to assessment items and made direct references to relying on their feelings and/or intuition to rationalize their response. In contrast, participants who broke the target attribute down into elements and examined each element took a decidedly more analytical and deliberate approach to their self-assessment.

Overall, the judgment strategies employed appeared to be cursory, particularly those used to generate initial item responses. System 1 thinking was pervasive, as evidenced by participants’ seemingly reliance on cognitive heuristics. For example, a total of eleven out of sixteen participants referred only to their dissertation research when responding to the self-assessment questions. Subsequent assessment items inquiring about the participants’ research experience revealed a number of other research projects that might have been used to surmise their research
efficacy; but only the most recent event was considered. This selective use of participants’ most recent example suggests the involvement of the availability heuristic in two ways, as discussed at length by Kahneman and Tversky (1972). First, a recency effect can be inferred given respondent’s use of the most recent research examples. Next, the availability heuristic can be used to explain respondent’s actual scores, which tended to be high. Kahneman and Tversky demonstrated that the easier it is for a participant to recall an example, the more important or more recent that example assumed to be (1972). In this case, it seems that importance ease of recall might have been used as a proxy for importance and competence, where participants used the ease of recall as evidence of competency on that attribute. One participant articulated this relationship, specifically:

“I can very easily think of a number of examples of me doing that for the work I did with [name of advisor] because it was a strength of mine.” (self-rating: 9/10 overall research skills)

The strategy of referring to a general pattern regarding the target attribute was also popular among participants in this study. This strategy was another example of Kahneman’s (2011) description of System 1 thinking, which favours associating new information with existing patterns or thoughts. Interestingly, participants leveraging this strategy as their primary judgment strategy were more likely to make reference to others’ perceptions to expand on their judgment later on. Arguably, this is another example of a mental shortcut, which might indicates the exclusive use of System 1 thinking. Moreover, some participants acknowledged that they were satisficing:

“I’m good with math so I’d be good at statistical analysis. My advisor has said many times that I’m a quick learner, so I feel confident that I would be able to learn statistical analysis and be good at it. That’s good enough for me!” (self-rating: 8/10)

Other judgment strategies, including referring to the consequences of an attribute, or
referring to a related trait, reflect Kahneman’s System 2 thinking. In the case of referring to consequences, the more deliberative thinking is evident in participants’ responses, which often included logical deductions based on recruited examples they used as evidence:

“My Master’s thesis went really well. I got invited to do various presentations about it, and my advisor was happy to keep working with me for my doctorate. I was even able to publish a couple of papers based on that work, you know, with my advisor type of thing. So I think that means I’m probably a pretty good researcher.” (self-rating: 8/10)

Similarly, participants who referred to a related trait engaged in a deliberate analysis of the trait, and explained their rationale for their choice of trait as a proxy for the target attribute.

“I’m very dedicated and I learn quickly. And also I’m good at reading articles and seeing new connections. I think these are the same things that would make me good at identifying different variables and coming up with what those variables could be, even if I haven’t actually done it yet.” (self-rating: 7/10)

Although it is possible that the generalized trait was not always a valid proxy, the deliberate, logical thinking participants engaged in support the activation of System 2 thinking.

Dual processing models have often treated the two systems as sequential, or at a minimum, mutually exclusive; while, Kahneman describes a more symbiotic relationship between the two (2011). In this view, System 1 is always “on”, and brings a variety of information to the (limited) attention of System 2, which rests in a minimally active state. A variety of triggers might engage System 2, causing it to override the answers generated by System 1, including the absence of knowledge, and surprise information (Kahneman, 2011). Additionally, the testing conditions, along with participants’ conscious and sub-conscious motivations, have been shown to affect the degree to which a hand-off from System 1 to System 2 occurs (Kahneman, 2011). For example, System 1 thinking is most likely to remain prevalent under conditions of stress, excessive cognitive load, or an environment with distractions (Kahneman, 2011). This seems reasonable given the cognitive demands of System 2 thinking on
attention, and consequently, the ease with which this thinking can be distracted (Kahneman, 2011).

The testing conditions for this study were optimal to encourage System 2 thinking. The interviews were conducted in a quiet room, only the interviewer and the participant were present. Interviews were scheduled for nearly double the length of time the assessment was expected to require. Yet, in spite of these ideal environmental conditions, the vast majority of participants defaulted to System 1 thinking; and even when System 2 thinking was engaged later during the assessment, participants’ responses were largely unaffected (they did not change from the initial response). This pattern prompts a need to examine further the possible variables affecting choice of judgment strategy, which represents the second main orientation in self-assessment and decision-making research. Schwarz, Hippler, Deutsch & Strack, (1985) noted that some respondents make systematic use of features of the questionnaire to arrive at an estimate. It is possible that the effectiveness of stating a question clearly and pointedly might affect the choice of judgment strategies and the System 1 and 2 thinking.

**Participant Knowledge**

Participant knowledge has been widely researched in the context of self-assessment. Most notably, the phenomenon of participants misunderstanding the measure, or lacking the knowledge to recognize their performance gaps has been discussed (Kruger & Dunning, 1999; Dunning, Heath & Suls, 2004). Kruger and Dunning (1999) refer to this phenomenon as being unskilled and unaware. In this scenario, participants with less knowledge tend to overestimate their performance; while participants with greater knowledge of the construct tend to be more accurate or to underestimate their performance on a particular attribute. This trend has been demonstrated repeatedly in a variety of contexts (e.g., Dunning, Johnson, Ehrlinger & Kruger,
and in each case, it has been assumed that the lack of knowledge, or reduced expertise renders the participants incapable of seeing their own gaps; while those with greater knowledge or superior expertise had a firmer grasp of their deficits.

Resulting from this research, it was expected that knowledge level, or in the case of this study, experience, would emerge as a key moderating variable in participants’ choice of judgment strategy. This appeared true in some instances, for example, participants who used generalized patterns, referred to others’ opinions, or relied on a feeling were more likely to be among the least experienced participants. These observations support Kruger and Dunning’s research, which demonstrated that participants’ lack of knowledge compromised their ability to recognize ‘what good looks like’; and consequently, the benchmark against which they could compare their own performance, was poorly defined. This resulted in participants lacking the detailed level of knowledge to understand the depth of their performance deficits. The current study provided an opportunity to explore this phenomenon from the participant’s perspective, and importantly, to invite participants to articulate the benchmark they used to evaluate their performance. Participant narratives for assessment scores suspected to be overestimations (because the participant was evaluating a skill they had never performed), and who employed non-specific judgment strategies (strategies that did not leverage specific examples of the target attribute) typically reflected a superficial understanding of the attribute.

“Oh so this is quantitative analysis right? So doing statistics….like…um…well using the software. Correlations. Yeah I think I’d be able to do that. I can learn. I just haven’t had to, but I would definitely be able to. So I would probably put myself at a 7-8 for that if I actually had to do it, I just don’t have to do it.” (self-rating: 7/10)

This superficial understanding is believed to have necessitated the selection of less specific judgment strategies in two ways: participants would be unable to generate relevant,
specific examples from memory, and would therefore need an alternative strategy for making their judgment; and participants’ lack of in-depth knowledge of the attribute would prevent them from recognizing when a strategy or the examples leveraged through their choice of judgment strategy were not adequate. These observations align with the prevailing research on the role of knowledge in self-assessment accuracy, as well as the research explicating the use of knowledge in forming judgments.

Interestingly, there was a large subset of the population (50%) who overestimated their abilities on skills they knew they did not possess. In other words, they systematically provided overestimated scores, while simultaneously acknowledging that they did not possess these skills, had never used these skills, and in some cases, did not understand all that might be encompassed by the skill. While this aligns with Kruger and Dunning’s (1999) general observations about the relationship between knowledge level and overestimation, it contrasts their hypothesis that participants are generally unaware of their deficiency.

An alternative explanation to the observations made for this subgroup draws upon Campbell and Sedikides’s research on self-threat and self-serving bias (1999). The motivation to protect, maintain, or enhance one’s positive self-concept has been widely reported (Brown & Dutton, 1995; Dunning, 1993; Sedikides, 1993; Sedikides & Strude, 1997). Threats to participants’ self-concept (e.g., via exposure to negative information), has been demonstrated to cause participants to take action to minimize or counter the self-threat (Campbell & Sedikides, 1999). The response patterns observed during this study support the presence of self-threat for most participants; and the subsequent triggering of a self-serving bias, which is related to the socially desirable response judgment strategy.

An analysis of the participants’ interviews demonstrated a clear belief about their abilities
as a researcher. Many participants referred to their experience and the very fact that they were in the doctoral program as evidence in support of their abilities as a researcher, which is akin to the reference to a specific event judgment strategy (Lindsay, refer to another strategy if you disagree):

“I’m really confident in my research abilities. I mean, I’m here, aren’t I? It’s a competitive program. And I’ve invested myself. I mean, this has been my life for the last almost five years. I’m a researcher. It even says that on my resume.”

According to Campbell and Sedikides, participants’ deliberate overestimation in confidence for skills they did not possess might be an attempt to preserve their self-concept as a good researcher (1999). The degree of overestimation (referred to as the absolute score and/or the articulated gaps in actual knowledge expressed by the participant) was higher when the attribute being assessed matched the research methodology the participant preferred (i.e., qualitative research skills prompted higher scores, even in the absence of the skill, for participants who identified as qualitative researchers). Interestingly, however, this observation was a matter of degree; target attributes that were distinctly part of the other research category continued to generate overestimated scores (i.e., qualitative researchers still overestimated their quantitative skills). It is possible that while participants’ identification as either a qualitative or quantitative research was important, the notion of being a good researcher overall was more critical to their self-concept, and thus more likely to evoke a self-serving response pattern (Campbell & Sedikides, 1999). This is also an example of the aforementioned “priming” psychological effect.

The use of knowledge, and by extension the availability of accurate knowledge, has been demonstrated to influence participants’ biases in other ways, in addition to the Dunning-Kruger effect. Slovic and Fischhoff, for example, demonstrated that the mechanism producing hindsight
bias resulted from participants’ failure to use their knowledge and powers of inference (1977). In other words, participants have been shown to make inaccurate assessments, or to apply inappropriate judgment strategies even when adequate knowledge is available to them. Although hindsight bias refers to the retrospective predictability of an event, similar mechanisms seem to be at work with this study’s participants. This might be true not only for participants who intentionally disregarded opposing evidence (i.e., delivered a socially desirable response to protect their ego), but also for those participants with a wealth of research experience, who continued to consider only one research project when reflecting on their skills.

The representativeness heuristic, which is triggered when participants are asked to judge probability under ambiguous circumstances, is another example of misuse of available knowledge or information. In the case of self-assessment, participants consider the ease with which examples are generated from memory to reflect the quality of their performance on the attribute being assessed. Conversely, Tversky and Kahneman demonstrated that the fact that an example is representative does not make it more likely; or said differently, the fact that a representative example can be generated from memory does not mean it is a behaviour the participants perform regularly, and with expertise.

The observations made during the cognitive interviews explicated some of the processes previously researched, and highlighted the conscious and sub-conscious processes at work in the selection of judgment strategies. By default, it appears most participants generate a generalized response, only later applying a more deliberate, analytical thought process to their evaluation. Level of knowledge and the presence of self-serving bias emerged as moderating variables in participants’ choice of judgment strategy. These interactive effects between cognitive processing, knowledge availability, and self-concept would benefit from more explicit qualitative
Q2&3: How does item grouping affect respondents’ judgment forming strategies, and how does item grouping affect respondents’ responses?

Item grouping is a special type of item order effect, which posits that clustering assessment items by category (e.g., qualitative research skills, quantitative research skills) elicits more reliable responses from participants (Lam, Green, & Bordignon, 2002). This improved reliability has been hypothesized to result from grouped items facilitating the adoption of a consistent mental set by the participant for each category of assessment items, thereby allowing them to leverage consistent points of reference in their evaluation.

Current research demonstrates mixed results in support of item grouping effects. While some research has demonstrated measurable improvement in reliability (Melnick, 1993; Lam, Green & Bourdignon, 2002), other studies have demonstrated negative affects on reliability (i.e., inflated scores) (Tourangeau & Rasinski, 1988; Hogarth & Einhorn, 1992); and others show no effect (Perlini, Lind & Zumbo, 1998). Apparently the research context in these different studies has some bearing on these contradictory findings.

This study was qualitative in nature, and thus cannot report on the internal consistency reliability of participants’ scores. In the absence of this quantitative support, however, this study explored patterns in judgment strategies and participant responses.

Overall, there were no notable trends in the selection of judgment strategies between the two conditions. Participants seemed equally likely to select an analytical approach, or to select a specific event to deconstruct for examples in the randomized condition as they were in the item-grouped condition. This can be observed in the near-equal distribution of judgment strategy types used in the two conditions. In the absence of differences in strategy choice, the two conditions did generate different patterns in participant responses (scores).
Differences between Non-Grouped Item-Grouped Conditions

Over half of the respondents in the randomized condition spontaneously attempted to categorize assessment items as either qualitative or quantitative after progressing through the first three to five assessment items. Interestingly, while this categorization did not appear to influence participants’ choice of judgment strategy, it did seem to influence the effort participants applied to leveraging the strategy. For example, referring to a specific event (dissertation research) was a commonly used judgment strategy in both conditions. Participants’ articulated recognition of the categorization of an item seemed to influence the degree to which that specific event was examined; relevant items (i.e., items matching the participants’ self-identified category) often drew more thoughtful analysis and more examples, while less-relevant items (i.e., items that did not match the participants’ self-identified category) were typically given a more cursory treatment,

“Oh I didn’t do that in my [dissertation] research so I can’t really evaluate that. I’d probably be a 6.” (self-rating: 6/10)

It seemed, therefore, that participants intuitively recognized some benefit to understanding the type of research skill they were assessing, as it provided signals as to the level of cognitive effort they should apply.

Two participants clarified this benefit,

“Oh so if this is about qualitative sample selection then yes I can do this because I had to do it. I know which examples to think about if you mean qualitative.”

“I think this is quantitative. Yes, it must be quantitative. Ok then I can tell you already I’m not good at it because I won’t have any examples. It makes me nervous just to think about it!”

This satisficing tendency is in-line with the research presented by Krosnick (1991), which highlighted human preference for minimal cognitive effort unless warranted. In this case,
participants provided what they considered to be a minimally sufficient response to the questions posed. Reflecting back on Kahneman’s two-system model of thinking, participants seemed to use the item categorization to justify System 1 thinking.

These results were somewhat unexpected, particularly with regards to participants’ spontaneous item categorization. Kahneman and Miller’s norm theory (1986) delivers a potential explanation for item grouping as follows: participants leveraged two general approaches to judging stimuli, a category-centered approach that compares a stimulus (or item) to the norms for a specific category; and a stimulus-centered approach that leverages norms (and examples) based on the specific elements of the item itself (Kahneman & Miller, 1986). It was expected that the randomized condition would force participants to take a stimulus-centered approach, thus affecting the type(s) of judgment strategies they would use.

In the item-grouped condition, references to the item categories were more explicit. Approximately 25% of participants in the randomized condition outright dismissed assessment items that were not in the category they self-identified with (in all cases, these were quantitative items),

“I don’t do quantitative research, so I can’t answer these. I’d be a zero, a nothing, because I don’t do them. That’s not to say I couldn’t, but I don’t so I can’t rate myself.”

While others prefaced those items with a disclaimer about their limited knowledge about these types of research activities.

“I’m not a quantitative researcher, so I haven’t done statistical analysis. So I wouldn’t expect myself to feel confident in these abilities since I don’t need them. But I’m good at learning so I’d probably be able to figure them out.”

Participants in the item-grouped condition also tended to share much more elaborate examples, expanding to include a level of detail not always observed in the randomized
condition. These observations are in-line with Kahneman and Miller’s norm theory, as “explicit reference to a category permits a high degree of control over the selection of norm elements. In particular, the use of a category label is very effective in restricting the evoked set to category members.” (Kahneman & Miller, 1986, pg. 141). Interestingly, however, these differences did not appear to result in significantly different choices in judgment strategies applied.

The most significant difference between the randomized and item-grouped conditions was observed in the scores participants assigned to themselves. Unlike in the randomized condition, many participants in the item-grouped condition were more willing to assign themselves significantly lower scores on items in the opposite category to the one they most identified with (i.e., scores of 1-3 out of 10). It is possible that the explicit categorization of assessment items lowered the self-threat experienced by these participants; they could excuse themselves for poorer performance on skills they deemed irrelevant or unnecessary to their work. Participants’ spontaneous categorization observed in the randomized condition did not seem to provide this same freedom (see Table 5).

Item grouping appeared to affect participants’ self-assessment by influencing their reported research efficacy. Specifically, the item-grouped condition appeared to generate more accurate scores for categories of research skills the participants acknowledged they did not possess. That is, participants in the item-grouped condition were significantly more likely to rate quantitative items lower than qualitative items. During the interviews, the non-grouped condition seemed to invite a self-threat related to participants’ self-perception as researchers, causing them to report inflated scores for skills they did not possess. Item grouping did not appear to influence the selection of judgment strategies, although the quality of examples generated in each conditioned differed, with the item-grouped condition enabling participants to recruit more specific, detailed
examples of relevant skills.

Specific use of the item category was not articulated in either condition, except in the rate attempt to clarify ambiguous or unknown terminology (e.g., “germane”). Participants did tend to draw upon the same specific event and examples, regardless of judgment strategy, throughout the assessment. It is unclear, however; whether this is indeed a function of the instrument, or an example of satisficing. Given the prevalence of this behaviour across conditions, it is likely that drawing upon the same example(s) is an illustration of satisficing, or, for some participants, the absence of alternative examples to refer to.

**Cognitive Interviewing**

The use of cognitive interviews, and specifically, a think-aloud protocol with follow-up questions proved effective in explicating the thinking behind participants’ self-assessment of research efficacy. As noted by a number of authors, the act of articulating thoughts in real-time inevitably affects a participants’ thinking (Ross, 1989). This was indeed observed during this study, begging the question as to whether and how participants’ judgment strategies were affected. For example, it is possible that the instruction to articulate their thoughts in real-time, and the desire to meet expectations by sharing something meaningful, might have encouraged participants to reflect and to use a more deliberate strategy (e.g., analysis, referring to a specific event) than they might normally do, which demonstrate the impacts of response format on the content and quality of participant responses (Hogarth & Einhorn, 1992; Krosnick and Alwin; 1987; Burton and Blair, 1991).

Nisbett and Wilson (1977) provided evidence that participants might have little or no direct access to their higher order cognitive processes. Instead, when subjects are required to report on these processes, they base their reports on implicit causal theories. This notion is
supported by cognitive research demonstrating that some aspects of the decision-making process simply are not accessible to the conscious mind. The steps participants take to retrieve and evaluate memories are automatic (System 1 thinking); and a number of studies have found that all participants have an immediate response to items that are not preceded by conscious reflection (Nisbett & Wilson, 1977; Taylor & Dionne, 2000; Slovic, Finucane, Peters & MacGregor, 2007). This observation was also made during this study; participants often articulated a score before sharing their rationale for the score.

Nisbett and Wilson’s research in this area was, however, at the same time reassuring. Participants have been demonstrated to generate accurate reports of their cognitive processes, even if these reports are, in part, context-driven (1977). Even in the absence of assured accuracy, therefore, this study is believed to have provided insights into the cognitive steps participants take during self-assessment. These insights have begun to support or challenge some of the observations and hypotheses expressed in current self-assessment research, and warrant further examination.
Chapter 6: Conclusion

This study attempted to explicate the cognitive processes behind the formation of self-assessment judgments, in the context of research efficacy. Current research in this area is limited to inferred processes based on the quantitative outcomes of assessments and their mediating and moderating variables. This research is therefore novel in its attempts to directly observe these inferred judgment processes. Additionally, this study attempted to explore the affects of item grouping on the judgment strategies used by participants to complete their self-assessment, and the effects of this grouping on self-assessment scores. Item grouping is another area in assessment research that has garnered little attention, despite being a strategy frequently employed (intentionally or not) by researchers.

The results of this study suggest there are eight judgment strategies typically employed by participants when forming self-assessment judgments. These strategies are not used in isolation, rather, appear to be used interchangeably, and seem heavily influenced by mediating variables. The mediating variables explored in this study included level of experience in the domain being evaluated (in this case, research skills/efficacy), self-identification as a qualitative or quantitative researcher, and item-grouping. Of these variables, item-grouping was the only significant variable influencing self-assessment scores; while experience in the domain and self-identification as a researcher appeared to have the strongest influence on participants’ choice of judgment strategy.

A number of self-assessment phenomenon that have previously been hypothesized through quantitative research were further explicated through this qualitative study. In some cases, the results of this study supported the current research, such as the illusory quality of initial judgments (Tversky & Kahneman, 1974; Hastie & Park, 1986; Stone, 1994). This study
demonstrated that while participants leveraged specific judgment strategies to solidify their assessment, nearly all participants seemed to formulate an immediate response without knowing where it came from. The judgment strategies were therefore used to rationalize their initial impression, or in some cases, to adjust them. Findings from this study corroborated other phenomena such as the role of expertise in self-assessment accuracy, and the significance of self-threat and the related self-serving bias. These phenomena jointly generated some surprising observations, even in the presence of expertise, self-serving bias can be triggered and override the need for accuracy. Increasing expertise, although a legitimate strategy for improving self-assessment accuracy in general, does not appear to mitigate the influence of self-threat and self-serving bias.

In addition to the observations made in-line with current research, a number of new or alternative observations were also made. These observations included a variation to the Dunning-Kruger effect, whereby participants overestimated their abilities even while recognizing their lack of skill in a given area. These observations beg the question of whether participants who lack expertise in an area are indeed “unskilled and unaware”, whether this scenario (not knowing what one is evaluating against) triggers a self-serving bias that causes them to overestimate their abilities (e.g., to cover for their lack of expertise and possibly save face), or are perhaps overconfident in their ability to gain a certain level of competency. Alternatively, this finding might be explained by the respondents’ inappropriate use of the rating scale and/or the ineffectiveness of the rating scale for measuring research efficacy. This study has no data to support this conjecture.

A particularly interesting observation was related to item-grouping. It was expected that the item-grouped condition would influence judgment strategies, and ultimately self-assessment
scores, by providing a consistent cognitive set within which to operate for each segment of the assessment. While grouping or not grouping items did not appear to generate different judgment strategies, the ungrouped condition introduced another behaviour that, itself, might be considered a supporting judgment strategy. Participants in this condition, almost unanimously, spontaneously categorized items as they progressed through the assessment. This categorization did seem to influence their judgments by focusing the examples they searched for and reported on, and in some cases, the score they assigned themselves.

The most pronounced difference between the item-grouped and ungrouped conditions was related to the role and impact of self-threat and the self-serving bias. Participants in the item-grouped condition were open to assigning themselves low scores on items for which they knew they did not possess the skill. Specifically, this was true for items that were in the opposite category to participants’ self-identification (i.e., qualitative researchers were willing to assign themselves low scores on quantitative items). Conversely, participants in the item-ungrouped condition seemed to resist assigning themselves truly low scores, even if they verbalized a lack of skill in the area being assessed. This occurred in spite of their categorization of the items. Participants’ self-identification of qualitative or quantitative researchers in this condition, therefore, was less important.

**Limitations and Future Directions for Research**

This study took a qualitative approach to exploring participants’ judgment strategies. While this approach provided new observations, the absence of quantitative data undermines reporting on the significance of these observations. This is particularly limiting when considering the moderating variables identified through these interviews. Future research should incorporate quantitative methods into the design, including measuring response time, strength of self-
identification as qualitative or quantitative researchers (i.e., the degree to which this label was an important part of their identity) and research expertise. Additionally, a mixed-method approach that paired cognitive interviews with a quantitative survey to obtain results from regression analyses between variables (e.g., expertise, self-threat, grouped versus ungrouped conditions) would provide deeper insights into the observations made here, which are limited to subjective patterns in the reported results. The use of qualitative analysis software might also provide a more robust level of analysis, particularly in a mixed-methods study.

This study was limited in scope to 16 participants, who were all doctoral students in the same department, and the majority of whom were female. The construct validity of participants is therefore limited by the degree to which the participants represent graduate students with qualitative and quantitative research training background. Future research should be conducted with a larger, more diverse (e.g., diversity of education, age, gender identity) sample size.

Research efficacy is a broad domain of knowledge leveraged specifically for this study due to the assurance of a common measure and minimum knowledge for participants. The scope of the domain explored in this study, however, was limited, which might explain the ease with which participants were able to categorize the items. This observation might have been compounded by the fact that participants had a relatively consistent level of knowledge about research skills (i.e., no participants were true beginners and no participants were true experts). Future research should consider using participants with a greater range of expertise within the domain, and assessment items that reflect greater complexity within the domain.

This study relied on participants’ think-aloud narratives to estimate participants’ self-assessment accuracy, which is rather subjective. Future research should include some objective criteria (such as performance on a short test of knowledge about qualitative and quantitative
research methods) to measure the accuracy of the self-assessment responses.

Subjectivity can also contaminate the accuracy of determining what judgment strategies are being reflected by the verbal reports. Some narratives can be linked to more than one strategy. Future research should further crystalize the description of the eight judgment formation strategies and the mechanism underlying these strategies. Furthermore, multiple researchers should be used to generate consensus about the strategies employed by the participants.

A final limitation of this study involves the use of cognitive interviewing as the primary research methodology. Although strong measures were taken to avoid affecting the data collection and analysis, it is possible that some aspects of the researcher’s personalities might have interacted with the data collection process. One aspect to consider is the interview protocol, perhaps the exercise of “thinking aloud” prompted participants to think and behave in ways that differ from typical self-assessment conditions. Moreover, as this study observed several obvious cases of self-serving bias, it is possible that the presence of the researcher and the request that participants verbalize their self-assessment scores introduced greater self-threat than would otherwise be present. Future research should explore variations to the methodology in an effort to draw out these biases.
Appendix A

Recruitment Email

Dear OISE doctoral students,

I am Lindsay Valve and I am a Masters of Arts student in the Curriculum, Teaching & Learning Department. Currently, I am conducting research on measuring doctoral students’ (who have completed at least one course in research methods) perception of their research efficacy. I am inviting you to participate in an interview-based self-assessment.

The interview is designed to examine the processes and strategies students use in responding to assessment items related to research efficacy. The interview will be conducted either on-line or in person at OISE, at a time that is convenient to you. The interview should take no more than 40 minutes.

If you wish to participate in either this research, please reply to this email to coordinate an interview time and location.

If you have any question about this research, you can email me at lindsay.valve@utoronto.ca

Thank you for your consideration to participate in my research.

Lindsay Valve
M.A. Student
Curriculum, Teaching & Learning Department

Recruitment Form

Graduate Student Research Self-Efficacy

You are invited to participate in a research study conducted by Lindsay Valve, who is a Masters of Arts student from the Department of Curriculum, Teaching, and Learning at the Ontario Institute of Studies in Education. Ms. Valve is conducting this study for her Master’s thesis. Dr. Tony Lam is her faculty supervisor for this project. This is an unfunded project.

Your participation in this study is entirely voluntary. You should read the information below and ask questions about anything you do not understand, before deciding whether or not to participate. You are being asked to participate in this study because you are a doctoral student at OISE, and as such, are likely to have completed at least one course in Research Methods.
• PURPOSE OF THE STUDY

The purpose of this study is to see how doctoral students feel about their research skills. We hope to use what we learn from the study to make changes to the program so it will help prepare doctoral students to conduct the research studies in support of their dissertations.

• PROCEDURES

If you volunteer to participate in this study, we will ask you to do the following:

1. We will ask you to take part in one face-to-face interview, lasting approximately one hour.
2. During this interview, you will be asked to complete a self-assessment on your research efficacy.
3. Upon completion of the self-assessment, the researcher will ask you to elaborate on some of your responses.
4. The interviews will be tape-recorded. These tapes will be destroyed once the information on them has been analyzed.

• POTENTIAL RISKS AND DISCOMFORTS

We expect that any risks, discomforts, or inconveniences will be minor and we believe that they are not likely to happen. If discomforts become a problem, you may discontinue your participation.

• POTENTIAL BENEFITS TO SUBJECTS AND/OR TO SOCIETY

It is not likely that you will benefit directly from participation in this study, but the research should help us learn how to improve the course offerings and other supports in place for students completing research at the doctoral level.

• COMPENSATION FOR PARTICIPATION

You will not receive any payment or other compensation for participation in this study. There is also no cost to you for participation, and the interview will take place at OISE to ensure its convenience.

• CONFIDENTIALITY

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law. Confidentiality will be maintained by means of a code number to let Ms. Valve and Dr. Lam know who you are. We will not use your name in any of the information we get from this study.
or in any of the research reports. When the study is finished, we will destroy the list that shows which code number goes with your name, and all of the interview recordings.

Information that can identify you individually will not be released to anyone outside the study. Ms. Valve will, however, use the information collected in her thesis and other publications. We also may use any information that we get from this study in any way we think is best for publication or education. Any information we use for publication will not identify you individually.

The tape-recordings that we make will not be heard by anyone outside the study unless we have you sign a separate permission form allowing us to use them. The tapes will be destroyed when Ms. Valve’s thesis is accepted.

In case of an emergency, injury, or illness that occurs during this study, I hereby authorize the release of any and all health information to allow for medical care and treatment of my condition.

• PARTICIPATION AND WITHDRAWAL

You can choose whether or not to be in this study. If you volunteer to be in this study, you may withdraw at any time without consequences of any kind. You may also refuse to answer any questions you do not want to answer. There is no penalty if you withdraw from the study and you will not lose any benefits to which you are otherwise entitled.

• IDENTIFICATION OF INVESTIGATORS

If you have any questions or concerns about the research, please feel free to contact

Ms. Lindsay Valve  
Principal Investigator  
Department of Curriculum, Teaching, and Learning  
University of Toronto  
Ontario Institute for Studies in Education  
Toronto, ON M5S 1V6  
416-220-1643  
lindsay.valve@utoronto.ca

Dr. Tony Lam  
Associate Professor  
Department of Curriculum, Teaching, and Learning  
University of Toronto  
Ontario Institute for Studies in Education  
Toronto, ON, M5S 1V6  
(416) 978-0140  
tcm.lam@utoronto.ca

• RIGHTS OF RESEARCH SUBJECTS

The University of Toronto Research Ethics Board has reviewed my request to conduct this project. If you have any concerns about your rights in this study, please contact Dr. Dean Sharpe of the University of Toronto Ethics Board at 416-978-5585.
I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

Printed Name of Subject

________________________
Signature of Subject
Date

Signature of Witness
Date
Appendix B

**Interview Script & Questions**

**Introduction**

Thank you for taking time to help us further explore doctoral students’ research efficacy at OISE. To provide you with more background on this project, doctoral students have been asked to participate in these interviews so that we can get a better understanding of the graduate experience regarding conducting research studies at OISE.

We have followed all of the human subjects protocols at OISE to carry out these interviews. As part of human subjects protection regulations, I will need you to sign a consent form.

There is plenty of time for the interview and I would like you to be as honest and thoughtful as possible in your responses and comments. For the first part of the interview, you will be asked to assess your research skills. As you complete this assessment, I will ask you to speak aloud about what you are thinking as you respond to questions. For example:

“Please tell me what you are thinking about when I ask you about your writing efficacy?”

“Please tell me what you are thinking about when I ask you about your ability to edit your own written work?”

After the assessment, I will ask you some follow up questions to better understand some of your responses.

Do you have any questions before we begin?

The following items are tasks related to performing research. Please indicate your ability to successfully perform each task by selecting a single number from one to ten that best describes your level of confidence. One would indicate no confidence, and ten would indicate extreme confidence. You can use these numbers or any of the numbers in between to describe your level of confidence. I would like to know how confident you are that you can successfully perform these tasks today. As you read and respond to these questions, please verbalize your thoughts. It is important that you are as honest and unfiltered as possible.

*Concurrent Verbalization Protocol Probes (if required):*

- What are you thinking?
- Tell me what you are thinking

*Retrospective Debrief Probes:*

- Tell me more about what you were thinking when
- How did you arrive at…
- Did you find the task difficult? Please explain…
Appendix C

Debriefing Form

Thank you for your participation in our study! Your participation is greatly appreciated.

Purpose of the Study:

We previously informed you that the purpose of the study was to examine the research efficacy of doctoral students at OISE. The goals of our research are to examine the cognitive strategies people use to evaluate their research skills, and to explore how item grouping (putting similar items together on an assessment) might affect the strategies people use to form judgments, and their responses.

Confidentiality:

You may decide that you do not want your data used in this research. If you would like your data removed from the study and permanently deleted please contact the principal investigator directly (Lindsay Valve, 416-220-1643, lindsay.valve@utoronto.ca). You may also inform us of your desire to withdraw at any time during your interview.

Final Report:

If you would like to receive a copy of the final report of this study (or a summary of the findings) when it is completed, please feel free to contact us.

Useful Contact Information:

If you have any questions or concerns regarding this study, its purpose or procedures, or if you have a research-related problem, please feel free to contact the researcher, Lindsay Valve, 416-220-1643 or email lindsay.valve@utoronto.ca.

If you have any questions concerning your rights as a research subject, you may contact the University of Toronto Research Ethics Board at 416-978-5585.

Further Reading(s):

If you would like to learn more about forming judgments during self-assessment or the item grouping effect please see the following references:


Lam, T.C.M., Green, K.E., Brodignon, C. (2002). Effects of item grouping and position of the “don’t know” option on questionnaire response. Field Methods, 14(4), 418-432.

***Please keep a copy of this form for your future reference. Once again, thank you for your participation in this study!***
Appendix D

**Background Questions**

<table>
<thead>
<tr>
<th>Items</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td>Male</td>
</tr>
<tr>
<td>Year of study at the doctoral level</td>
<td>1  2  3  4  5  6  7  8</td>
</tr>
<tr>
<td>How many research methods courses have you taken at OISE?</td>
<td>1-2</td>
</tr>
<tr>
<td>What research methods courses have you taken at other institutions?</td>
<td>1-2</td>
</tr>
<tr>
<td>How many research studies have you participated in as a subject?</td>
<td></td>
</tr>
<tr>
<td>How many research studies have you participated in as a researcher?</td>
<td></td>
</tr>
<tr>
<td>How confident are you in your research skills?</td>
<td>Not</td>
</tr>
<tr>
<td></td>
<td>Confident</td>
</tr>
<tr>
<td></td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>Would you consider yourself a qualitative or a quantitative researcher?</td>
<td></td>
</tr>
</tbody>
</table>

**Non-Grouped Condition (Form A)**

<table>
<thead>
<tr>
<th>Items</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>I can develop an interview script and questions</td>
<td>Not</td>
</tr>
<tr>
<td></td>
<td>Confident</td>
</tr>
<tr>
<td></td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>I can explain the concept of validity of measurements</td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>I can use software to conduct statistical analysis</td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>I am able to identify dependent, independent, and extraneous variables in designing a research study</td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>I am able to select a participant sample that is representative of a population by using established sampling procedures</td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>I am able to set up and test hypotheses using numerical data</td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>I can take scrupulous notes</td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>I am able to code data from recorded interviews</td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>I am able to purposefully identify individuals germane to my research topic from whom to gather data</td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>I can observe participants using open, unstructured recording</td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
</tbody>
</table>
How confident are you in your research skills? | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10
--|---|---|---|---|---|---|---|---|---|---
How confident were you in your research skills just before you joined OISE? | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10

### Item-Grouped Condition (Form B)

<table>
<thead>
<tr>
<th>Items</th>
<th>Not Confident</th>
<th>Extremely Confident</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative research is an approach to acquiring knowledge through the in-depth study of a small number of cases in natural environments and with narrative data.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I can develop an interview script and questions</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I can observe participants using open, unstructured recording</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I am able to code data from recorded interviews</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I am able to purposefully identify individuals germane to my research topic from whom to gather data</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I can take scrupulous notes</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

**Quantitative research is an approach to acquiring knowledge through the manipulation of treatment or by retrospectively analyzing treatment effects with a large number of participants. The data are numbers and statistics.**

<table>
<thead>
<tr>
<th>Items</th>
<th>Not Confident</th>
<th>Extremely Confident</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am able to set up and test hypotheses using numerical data</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I am able to select a participant sample that is representative of a population by using established sampling procedures</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I can use software to conduct statistical analysis</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I am able to identify dependent, independent, and extraneous variables in designing a research study</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I can explain the concept of validity of measurements</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>How confident are you in your research skills?</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>How confident were you in your research skills just before you joined OISE?</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
References


Lam, T.C.M., Green, K.E., Brodignon, C. (2002). Effects of item grouping and position of the “don’t know” option on questionnaire response. *Field Methods, 14*(4), 418-432.


