University of Louisville

ThinkIR: The University of Louisville's Institutional Repository

Electronic Theses and Dissertations

12-2011

Exploring the scholar-practitioner gap in personnel selection assessments : an analysis of scholarly versus practitioner literature.

Whitney L. Martin 1977-
University of Louisville

Follow this and additional works at: <https://ir.library.louisville.edu/etd>

Recommended Citation

Martin, Whitney L. 1977-, "Exploring the scholar-practitioner gap in personnel selection assessments : an analysis of scholarly versus practitioner literature." (2011). *Electronic Theses and Dissertations*. Paper 908.

<https://doi.org/10.18297/etd/908>

This Master's Thesis is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact thinkir@louisville.edu.

**EXPLORING THE SCHOLAR-PRACTITIONER GAP IN
PERSONNEL SELECTION ASSESSMENTS:
AN ANALYSIS OF SCHOLARLY VERSUS PRACTITIONER LITERATURE**

By

**Whitney L. Martin
B.A., University of Kentucky, 1999**

**A Thesis
Submitted to the Faculty of the
College of Education of the University of Louisville
In Partial Fulfillment of the Requirements
For the Degree of**

Master of Science

**Department of Leadership, Foundations & Human Resource Education
University of Louisville
Louisville, Kentucky**

December 2011

EXPLORING THE SCHOLAR-PRACTITIONER GAP IN
PERSONNEL SELECTION ASSESSMENTS:
AN ANALYSIS OF SCHOLARLY VERSUS PRACTITIONER LITERATURE

By

Whitney L. Martin
B.A., University of Kentucky, 1999

A Thesis Approved on

November 18, 2011

By the following Thesis Committee:

A. Herd (Thesis Director)

Jill L. Adelson

ABSTRACT

EXPLORING THE SCHOLAR-PRACTITIONER GAP IN PERSONNEL SELECTION ASSESSMENTS: AN ANALYSIS OF SCHOLARLY VERSUS PRACTITIONER LITERATURE

Whitney L. Martin

November 18, 2011

Research suggests that a gap exists between scholarly findings and practitioner knowledge, beliefs, and practices in the Human Resource field, particularly in the area of employee selection (Deadrick & Gibson, 2007; Rynes, Giluk, & Brown, 2007). This study seeks to explore this gap relative to self-report selection assessments by examining practitioner-oriented versus scholarly literature. Articles published between January 2006 and September 2011 from two scholarly sources (*Journal of Applied Psychology* and *Personnel Psychology*) and two practitioner sources (*HR Magazine* and *HR Executive*) were reviewed, and 49 articles were selected for inclusion in analysis. Qualitative content analysis was used to analyze the articles relative to five themes: purpose of the article, type of selection assessment discussed, specific instruments mentioned, how validity was discussed, and how utility was discussed. It was found that there were significant differences in the way that scholarly and practitioner publications discussed assessments, especially in the areas of validity and utility. Implications for scholars and practitioners are discussed.

TABLE OF CONTENTS

	PAGE
ABSTRACT.....	iii
LIST OF TABLES.....	vi
 CHAPTER 1: INTRODUCTION.....	 1
Significance of Study.....	5
Scope and Research Questions	5
 CHAPTER 2: LITERATURE REVIEW.....	 14
Scholar-Practitioner Gap.....	15
General Mental Ability.....	18
Personality.....	21
Emotional Intelligence.....	25
Integrity.....	29
Summary of Literature Review.....	32
 CHAPTER 3: METHODOLOGY.....	 36
Scholar and Practitioner Journal Selection	36
Timeframe.....	37
Procedure.....	38
Content Analysis.....	40

CHAPTER 4: RESULTS.....	42
Theme 1.....	42
Theme 2.....	45
Theme 3.....	46
Theme 4.....	47
Theme 5.....	50
Summary.....	53
CHAPTER 5: DISCUSSION.....	55
Theme 1.....	55
Theme 2.....	56
Theme 3.....	57
Theme 4.....	59
Theme 5.....	60
Implications for Practitioners.....	61
Implications for Scholars.....	63
Limitations of Current Study.....	64
Suggestions for Future Research.....	65
Conclusion.....	66
REFERENCES.....	68
APPENDIX.....	79
CURRICULUM VITAE.....	103

LIST OF TABLES

TABLE	PAGE
1. Descriptions of Common Terms Used in this Study.....	10
2. Predictive Validities of Various Hiring Methodologies.....	15
3. Number of Articles Included from Each Source.....	40
4. Purpose of Articles by Source.....	44
5. Category of Assessment Referenced by Source.....	45
6. Number of Articles per Validity Category by Source.....	50
7. Number of Articles per Utility Category by Source.....	53

CHAPTER 1

INTRODUCTION

Research on the utility of using valid selection systems leaves little doubt that getting the right people into the right organizations and the right jobs can make a big difference. Popular business publications have delivered a similar message regarding selection—in the best seller Good to Great, for example, Collins (2001) wrote about the importance of “getting the right people on the bus”—yet most practitioners still aren’t aware of some of the most important findings from selection research (Rynes, Giluk, & Brown, 2007, p. 1001).

Well-developed assessment tools can help companies consistently hire better performing employees. Thousands of studies have been conducted in academia and industry surrounding the psychology of individual differences in intelligence, various aptitudes and abilities, personality, interest, values, and a myriad of other traits that systematically affect job performance (Schmidt & Hunter, 1998; Schmidt & Hunter, 2004). Assessment instruments have been developed to measure these traits and abilities reliably and have been proven to be valid predictors of performance across a wide range of jobs and settings (Murphy, 2000). As a result, employers have at their disposal valid, reliable, and relatively low cost selection methods that can substantially reduce selection errors and increase the productivity, output, and effectiveness of their workforces (Schmidt & Hunter, 1998; Murphy, 2000; Scroggins, Thomas, & Morris, 2008).

However, studies suggest that these findings based on scholarly research have not effectively transferred to Human Resources (HR) practitioner knowledge or practice (Rynes, Colbert, & Brown, 2002). In their 2002 study, Rynes and colleagues surveyed

nearly 1000 HR practitioners to determine the extent to which practitioners knew about and believed 35 well-established scholarly research findings. The area of greatest disconnect was in “staffing,” where less than half of the practitioners agreed with prevailing research findings. Specifically, practitioner responses indicated they did not believe that intelligence is a better predictor of performance than personality or values or that integrity tests can effectively predict counterproductive workplace behaviors. Practitioner responses also indicated their belief that there are four basic personality dimensions (as represented in the Meyers Briggs Type Indicator) and that there is little difference between various personality inventories in their ability to predict job performance.

In contrast to the four above stated practitioner beliefs, it is a foundational truth in selection research that general mental ability (GMA) is one of the most effective predictors of employee performance (Christiansen, Janovics, & Siers, 2010; Furnham, Dissou, Sloan, & Chamorro-Premuzie, 2007; Schmidt & Hunter, 1998; Schmidt & Hunter, 2004; Terpstra & Rozell, 1993; Wanek, 1999). Research has also suggested that professionally developed integrity tests are valid predictors of both productive and counterproductive behaviors (Ones, Viswesvaran, & Schmidt, 1993; Wanek, 1999), with an average validity of $r = .41$ in predicting supervisory ratings of overall job performance (Ones et al., 1993; Schmidt & Hunter, 1998). In addition, most researchers have generally come to accept the Five Factor Model (not four-factor) as a universal, all-encompassing model of personality structure (Furnham et al., 2007; Judge, Martocchio, & Thorensen 1997). Lastly, there are wide variations in the reliability and accuracy of

scores produced from various personality assessment instruments (Barrick & Mount, 1991).

In addition to their lack of *knowledge* about some of the most effective selection methods, some studies have also found that HR practitioners opt not to rely on the most valid predictors of performance in their selection *practices*. There appears to be widespread reliance on selection practices not well supported by empirical research, such as handwriting analysis and unstructured interviews, whereas those practices that have been proven more effective, such as personality, honesty-integrity, and cognitive ability test, are used less consistently (Anderson, Lievens, van Dam, & Ryan, 2004). In a recent study, Piotrowski and Armstrong (2006) found that only about 20% of employers use personality tests and only 28% screen for honesty-integrity. In addition, a survey by Terpstra and Rozell (1997) found that only 20% of companies reported using cognitive ability tests. These findings point toward the conclusion that there remains a substantial gap between what academic literature says and what most practitioners actually do (König, Klehe, Berchtold, & Kleinmann, 2010; Murphy, 2000; Terpstra & Rozell, 1993).

However, even if HR practitioners knew about the effectiveness of GMA and various personality/integrity constructs in predicting performance, translating that knowledge into practice may be a challenge for the following reasons. Even after the business need for implementing a more stringent selection process has been identified, the practitioner would need to decide what constructs to measure and would need to identify available tools that do so in a valid and reliable manner, all while analyzing cost-benefit ratios, administrative requirements, potential legal concerns, etc.—a daunting task. Furthermore, the marketplace is virtually saturated with assessment instruments—of

varying degrees of quality (Scroggins et al., 2008)—and human resources professionals are often faced with the task of separating the chaff from the wheat, so-to-speak. In addition, questionable sales tactics (Sackett, Burris, & Callahan, 1989) and publication bias by test publishers make clearly and accurately reported validation data the exception rather than the rule (McDaniel, Rothstein, & Whetzel, 2006). Furthermore, many HR practitioners may not be trained in statistical analysis or psychometrics necessary for a thorough understanding and critical review of assessment instrument construction and validation data (Rynes et al., 2002; Terpstra & Rozell, 1997). Considering all this, it is perhaps easy to understand why so many practitioners have decided to continue relying on selection practices that may be less effective but with which they are more comfortable.

Where might a human resource practitioner turn to acquire the information necessary to become a shrewd and informed consumer of selection assessments? One study of HR practitioner behavior found that only a negligible percentage (<1%) of HR practitioners read scholarly journals (Rynes et al., 2002). Therefore, the purpose of the present study is to review the publications that HR practitioners *do* read to better understand what information practitioners receive regarding selection assessments that shape their beliefs and inform their practices. Scholarly journals geared toward general HR issues are also analyzed from the same time period to determine the relative emphasis placed on selection assessments in these publications and to assess the focus and content of the relevant studies. In this way, it can be ascertained to what extent there is consistency between the information being published by scholars and the information being disseminated to practitioners. To the extent that the content from the two sources

differ, insight can be gained about potential reasons for the scholar-practitioner “gap” in personnel selection assessment beliefs and practices.

Significance of Study

The current study seeks to further research on the scholar-practitioner gap by investigating the extent to which scholarly and practitioner journals are aligned in their coverage of workplace selection assessments. As those in the human resource profession seek to become more strategic (Terpstra & Rozell, 1993) and have a greater impact on bottom-line business outcomes, the decisions they make relative to their company’s hiring process are critical (Terpstra & Rozell, 1997). If HR practitioners are opting not to use selection methods with high predictive validity, either due to a lack of knowledge or erroneous beliefs about utility, the negative outcomes will include higher turnover, poor job fit, lower employee engagement, and impact to bottom-line organizational performance (Rynes et al., 2002; Rynes, et al., 2007; Sackett et al., 1989; Schmidt & Hunter, 1998; Terpstra & Rozell, 1993; Terpstra & Rozell, 1997).

Scope & Research Questions

This study focuses on *commercially available paper-and-pencil (or internet based) self-report tests/inventories/instruments* used for employee *selection*. This topic is important for study for numerous reasons. First, the Rynes et al. (2002) study found that the greatest discrepancy between research findings and practitioner knowledge and beliefs existed in the area of “staffing/selection.” The preponderance of the items in that category concerned intelligence and personality tests (of which honesty/integrity tests are a subset) as predictors of employee performance. The Rynes et al. (2007) study then confirmed that these are among the most *important* research findings for HR practitioners

to understand. Furthermore, cognitive tests have been shown to be the strongest predictor of job performance, and personality tests have been one of the most extensively and widely studied (Van Rooy, Viswesvaran, & Pluta, 2005; Van Iddekinge, Putka, & Campbell, 2010; Komar, Brown, Komar, & Robie, 2008). Therefore, this study is primarily interested in how intelligence tests and various types of personality-based tests used for employee selection have been addressed in practitioner and scholarly literature.

Paper-and-pencil tests of GMA and personality are very well defined in terms of what constructs they measure and with what degree of accuracy—"well developed self report questionnaires serve as useful measures of personality due to their sound psychometric properties, their rapid assessment of numerous job-relevant traits, and their ability to predict various dimensions of job performance" (Bing, Stewart, Davison, Green, McIntyre, & James, 2007, p. 722). And, the psychometric properties and criterion-related validity of professionally developed, commercially available validated instruments have been subjected to rigorous scrutiny (Ones, Dilchert, Viswesvaran, & Judge, 2007; Tett & Christiansen, 2007). So, from a practical standpoint, if a practitioner were to learn of the research findings concerning the predictive validity of GMA and personality constructs, a very clear and easy strategy they could implement would be to integrate a commercially available instrument scientifically designed to measure the specific construct(s) of interest. The question then becomes, have practitioners been educated, through practitioner literature, on what criteria such tests should meet, particularly in the areas of validity and utility, in order to be shrewd consumers of these products? This study will seek to answer this question.

Next, this study is interested in how publications communicate information about assessment instruments in a *selection* context. Assessments can be helpful in many business contexts, including teambuilding, coaching, succession planning, and employee development. However, different measures are more relevant and effective in different contexts, which explains why studies (e.g.: Christiansen et al., 2010; Schmidt & Hunter, 1998; Van Rooy et al., 2005) often delineate between a construct's or instrument's ability to predict job performance versus performance in a training class (as an example) or other context. Furthermore, many of the concerns HR managers express surrounding the use of assessments are specific to the selection context (i.e., adverse impact). For these reasons, this study focuses on articles that speak to the use of assessments for selection purposes, rather than in the context of other business endeavors (like training). Many other studies have also limited their focus to a selection context (e.g., Aronson & Reilly, 2006; Arthur, Bell, Villado, & Doverspike, 2006; Christiansen et al., 2010; Heggstad, Morrison, Reeve, & McCloy, 2006; Sackett et al., 1989; Tsaousis & Nickolaou, 2001; Wanek, 1999).

Based on a review of the literature, the purpose of this study is to investigate the scholar-practitioner gap in the area of selection assessment instruments. In reviewing practitioner versus scholarly publications, selected articles will be analyzed for content relevant to five specific research questions. The questions and the rationale for inclusion of each question follow:

1. *What is the main purpose of the article?*

Coding each article as to its primary message/purpose will allow the degree of consistency between practitioner and scholarly publications in their focus as it

relates to selection assessments to be determined. Deadrick & Gibson (2007) used this method in their study to analyze the “interest gap” between academicians and practitioners. This study has a similar goal within the narrower scope of employee selection instruments.

2. *What category of workplace selection assessment is discussed?*

By categorizing and counting the references to various types of selection assessments (e.g.: intelligence/aptitude, personality, honesty/integrity, etc.) it will be possible to determine whether HR practitioners are receiving information about the same types of constructs and measures that scholars are researching. For example, in the Rynes et al. (2007) study, there was found to be a large gap in the area of GMA or intelligence testing in that while this construct is widely studied in academic circles, zero articles appeared in *HR Magazine* (the practitioner magazine they reviewed) during the timeframe of their study.

3. *Are any commercially available assessment instruments mentioned specifically? If so, which ones?*

In a preliminary study of scholarly articles relative to selection assessments, the primary researcher noticed that very few scholarly articles tended to discuss specific instruments that are commonly used in organizations in a pre-employment context. Conversely, many widely used tools in industry do not seem to have been reviewed or analyzed by scholars. Analysis of articles from both scholarly and practitioner sources will enable a quantitative assessment of the degree to which academics and practitioners have experience with the same instruments.

4. *How is validity discussed (e.g., What language is used to discuss validity? How is validity reported?)*

In order for practitioners to select a pre-employment assessment instrument, they have to possess some knowledge regarding the existence of and importance of validity, as well as how the validity of an instrument is determined. Analyzing how validity is discussed and reported in scholarly versus practitioner literature will allow the extent to which both groups are “speaking the same language” to be determined and, to the extent that they are not speaking the same language, may provide insight into a reason for the reported “knowledge gap” relative to selection assessments.

5. *To what extent and in what manner does the article address the concept of utility (i.e., the expected benefit from using assessments)?*

Utility can be defined as the “practical economic value” of using selection assessments (Schmidt & Hunter, 1998, p.262). In examining utility, how each source discusses the benefits of using selection assessments will be analyzed to determine the extent to which scholars and practitioners share similar rationales and justifications for using assessment instruments in a selection context (see Table 1 for further discussion of “utility” and other common terms used in this study).

Table 1

Descriptions of Common Terms Used in this Study

Common Term	Description
Reliability	Reliability refers to the consistency or dependability with which a test measures some construct without being affected by random error. Two commonly reported measures of reliability are coefficient alpha, which represents the inter-relatedness and internal consistency of test items, and test-retest reliability which demonstrates the degree to which the trait being measured is stable and not easily influenced by temporary changes in people at the time of testing. Reliability coefficients range from 0 to 1 and are commonly reported as a decimal, with higher numbers indicating higher reliability. In their manual "Testing and Assessment: An Employer's Guide to Good Practices" (1999), the U.S. Department of Labor suggests that reliability coefficients in the .7 range are adequate, in the .8 range are good, and in the .9 range are excellent.
Validity	Validity refers to the effectiveness with which an instrument measures what it purports to measure. Two types of validity often analyzed in the context of selection assessments are criterion-related and construct validity. Criterion-related validity measures the relationship between test scores and the outcome-of-interest. For example, if a test is designed to measure employee absence and candidates who do well on the test subsequently have a better attendance record than those

scoring poorly on the test, the test would be seen to have criterion-related validity. Criterion-related validity can be assessed through concurrent or predictive validation studies. Construct validity reflects how well the instrument actually measures the construct or characteristic that it claims to measure and not some other construct. A validity coefficient is also expressed as a decimal between 0 and 1, with higher numbers indicating a greater degree of confidence one can have in the prediction made by the test. However, because job performance is affected by so many different factors, an individual test can only be expected to predict a certain percentage of the variance. Therefore, it is unusual to find validity coefficients (in the social sciences) that exceed .4 (U.S. Department of Labor, 1999) and validities in the range of .21-.35 are considered acceptable for a single test (Scroggins et al., 2008).

Incremental
Validity

Incremental Validity can be achieved by combining two measurement methods that are not correlated with each other and do not overlap in what they measure. When this happens, addition of a second measure offers incremental validity, or additional predictive value beyond what was obtained by using the first measure by itself. This is important because an employer could use multiple strategies, like a cognitive ability test and an assessment center, but because a large portion of what is measured in an assessment center is general mental ability, the assessment center, though a measure that allows for valid inferences in

and of itself, does not offer substantial incremental validity beyond what is attained by using the cognitive test alone.

Meta-analysis	Reliability and validity estimates in individual studies can be affected by statistical and measurement artifacts like sampling error or measurement error of certain constructs (like job performance), which reduce statistical power. One way to counteract this is by conducting a meta-analysis. A meta-analysis essentially synthesizes findings from a large number of studies, combining validity estimates and correcting for statistical errors due to sampling errors, measurement errors, etc. so that the findings can be better generalized across settings.
Job Performance	Many assessments purport to measure things that affect future job performance. In this case, it is critical that there be clarity around how the construct of “job performance” is operationalized. In most of the studies referenced herein, job performance was assessed through manager ratings, though some studies also looked at production records, sales records, or other indicators of job-related outcomes. Because there tends to be a great deal of variability in performance and output among workers in most situations, using a selection tool that can predict which applicants will be most productive and effective is very important.
Utility	Utility speaks to the “practical value” (Schmidt et al., 2008, p.262) of using a selection instrument. Utility is usually measured by output, dollars, mean performance, etc. (Berry, Clark, McClure, 2011). The

economic value added by a selection method is determined by the predictive validity of the tool relative to the cost of implementing it (Mount et al., 2008) as well as a variety of other factors including its usefulness and value in the context of the organization's overall selection procedures (Morgeson, Campion, Dipboye, Hollenbeck, Murphy, & Schmitt, 2007b).

CHAPTER 2

REVIEW OF LITERATURE

If it were the case that all job candidates would perform equally well in a given position, there would be no need for selection devices of any kind. However, because of the wide variation in worker performance levels across job types, it is critical for organizations to understand what differences among individuals systematically affect job performance so that the candidates with the greatest probability of success can be selected (Schmidt & Hunter, 1998; Schmidt & Hunter, 2004). Schmidt and Hunter (1998) conducted a meta-analysis of 85 years of research on selection practices and were able to determine the mean predictive validity of 19 selection procedures for predicting job performance (as determined by supervisor ratings, production records, sales records and other measures). Validity estimates were corrected for downward bias due to measurement error in the measures of job performance. Results of particular interest to this study are presented in Table 2.

It is interesting to compare the validities of some of the most effective measures (i.e.: GMA, $r = .51$, and integrity Tests, $r = .41$) to the validities of the practices most commonly used by organizations, like unstructured interviews ($r = .38$) and reference checks ($r = .26$, which may be an over-estimation of the validity considering the current legal climate in the United States and employers' reluctance to share potentially damaging information about past employees) (Furnham, 2008; Schmidt & Hunter, 1998). Schmidt & Hunter (1998) concluded that organizations that choose to rely on less predictive selection methods are unnecessarily creating a competitive disadvantage for

themselves. This chapter provides an overview of the research concerning the science-practitioner gap and a review of the research relative to the most common types of self report assessments—GMA and personality (which includes emotional intelligence and integrity tests). This chapter concludes with the research questions that guide this study.

Table 2

Predictive Validities of Various Hiring Methodologies

Selection Method	Validity	Incremental
		Validity over GMA
General Mental Ability (GMA)	.51	n/a
Integrity Tests	.41	.27
Assessment Centers	.37	.04
Conscientiousness Tests	.31	.18
Interest Tests	.10	.02
Graphology	.02	0

Note. Modified from the table found on p. 265 of Schmidt & Hunter (1998).

The Scholar-Practitioner Gap

The research-practice gap in human resources has been widely acknowledge and documented. In their 2002 study, Rynes and colleagues surveyed nearly 1000 HR practitioners to determine the extent to which the practitioners knew about and believed 35 well-established scholarly research findings. The area of greatest disconnect was in “staffing,” where less than half of the practitioners agreed with prevailing research findings on all but one item. Specifically, practitioner responses indicated they did not

believe that intelligence is a better predictor of performance than personality or values or that integrity tests can effectively predict counterproductive workplace behaviors. Practitioner responses also indicated the belief that there are four basic personality dimensions (as represented in the Meyers Briggs Type Indicator) and that there is little difference between various personality inventories in their ability to predict job performance.

In contrast, it is well established in selection research that (GMA) is one of the strongest predictors of employee performance (Christiansen et al., 2010; Furnham et al., 2007; Schmidt & Hunter, 1998; Schmidt & Hunter, 2004; Terpstra & Rozell, 1993; Wanek, 1999). Thousands of validity studies have been conducted on GMA tests, with meta analyses demonstrating that cognitive tests exhibit validity across various jobs and settings and are not bound by situation specificity as was once thought (Schmidt & Hunter, 2004). Research has also shown that professionally developed integrity tests are valid predictors of counterproductive behaviors (Ones et al., 1993; Wanek, 1999), with an average validity of $r = .41$ in predicting supervisory ratings of overall job performance (Ones, Viswesvaran, & Schmidt, 1993; Schmidt & Hunter, 1998). In addition, most researchers have generally come to accept the Five Factor Model (not four-factor) as a universal, all-encompassing model of personality structure (Furnham et al., 2007; Judge, Martocchio, & Thorensen 1997).

These erroneous beliefs related to selection assessments appear to translate into practice. It often is found to be the case that those HR selection practices not well supported by empirical research, such as handwriting analysis and unstructured interviews, remain popular, whereas those that have been proven more effective, such as

personality, honesty-integrity, and cognitive ability tests, are used less consistently (Anderson, Lievens, van Dam, & Ryan, 2004). In a recent survey of 750 companies randomly selected from national business directories, of which 151 firms participated, HR practitioners Piotrowski and Armstrong (2006) found that only about 20% of employers use personality tests and only 28% screen for honesty-integrity. In addition, a survey study by Terpstra and Rozell (1997) found that 20% of companies reported using cognitive ability tests.

In trying to diagnose potential causes of the disconnect between academic and practitioner beliefs and practices, one area of study has been the discrepancies in what is reported in scholarly journals as opposed to practitioner literature. Identifying gaps in the literature read by HR scholars and practitioners may suggest that a research-knowledge gap is fueling the research-practice gap. In their 2007 *Academy of Management Journal* article, "The Very Separate Worlds of Academic and Practitioner Periodicals in Human Resource Management: Implications for Evidence-Based Management," Rynes and colleagues explored the research-practice gap by analyzing the content of *HR Magazine*, *HR Management*, and *Harvard Business Review* relative to three known practice gap areas: the importance of intelligence or GMA for performance, the importance of goal setting and feedback for performance, and the validity of personality and integrity tests for predicting performance. Their findings confirmed "the significant failure of academic research to transfer to important practitioner sources of information" (p. 999). Not only were these important topics sparsely addressed (if at all) in practitioner literature, but the messages conveyed and evidence cited were not generally consistent with scholarly findings.

This study specifically investigates the gaps between scholars and practitioners in the area of HR selection. The area of HR selection is particularly critical given the nature of economic and global changes affecting organizations and employees in recent years (Terpstra & Rozell, 1997). Organizational selection practices for recruiting, staffing, and identifying top talent in order to make sound promotion, development, and retention decisions affect all important HR outcomes, including individual, team, and organizational performance, employee satisfaction and perceptions of equity, employee turnover, and talent management, to name a few (Heneman & Judge, 2008; Rynes et al., 2002). One method practitioners can use to increase the rigor of their selection practices is to implement a commercially available self-report assessment of GMA and/or personality. Such instruments are useful “due to their sound psychometric properties, their rapid assessment of numerous job-relevant traits, and their ability to predict various dimensions of job performance” (Bing et al., 2007, p. 722). An overview of the scholarly research surrounding GMA and personality assessments follows.

General Mental Ability (GMA) Assessments

Research has firmly established that GMA is the best predictor of employee performance (Christiansen et al., 2010; Furnham et al., 2007; Schmidt & Hunter, 1998; Schmidt & Hunter, 2004; Terpstra & Rozell, 1993; Wanek, 1999). Thousands of validity studies have been conducted on GMA tests and meta analyses have demonstrated that cognitive tests have validity across various jobs and settings and are not bound by situation specificity as was once thought (Schmidt & Hunter, 1998; Schmidt & Hunter, 2004). It has also been found that as jobs become more complex, the validity of GMA

tests increase (Murphy, 2000; Ones, Viswesvaran, & Schmidt, 1993; Schmidt & Hunter 1998; Schmidt & Hunter, 2004).

GMA tests have been in use since the end of World War I and typically contain questions related to verbal, numerical, spatial, and reasoning ability (Schmidt & Hunter, 2004). GMA can be conceptualized as a hierarchy with general intelligence at the top, broad content abilities (i.e., verbal, numerical, and spatial) in the middle, and specific aptitudes such as spelling and paragraph comprehension, at the bottom (Mount et al., 2008). Among GMA tests, there are different types of intelligence that can be assessed. Two primary categories of interest are fluid intelligence and crystallized intelligence. Fluid intelligence includes information-processing and reasoning ability. Crystallized intelligence refers to the ability to acquire, retain, organize, and conceptualize information (Furnham et al., 2007). Some of the most widely accepted GMA tests, according to the academic literature, are the Graduate Management Assessment (GMA), the Watson-Glaser Critical Thinking Appraisal (WGCTA), and Wonderlic assessments (Furnham et al., 2007; Schmidt & Hunter, 2004).

One reason that GMA is an effective predictor of job performance is because higher GMA allows a person to acquire job knowledge faster, which translates into better performance (Schmidt & Hunter, 1998). However, this is not only applicable when a person initially begins a new role. The relationship between GMA and job performance remains stable throughout an employee's tenure in a position (Schmidt & Hunter, 2004). This is likely because, in today's world, jobs change rapidly and require constant learning, adaptation, and innovation. Because GMA has been found to correlate highly with divergent thinking and problem-solving abilities (Anderson et al., 2004), it will

continue to become increasingly important as the rate of change in organizations becomes even more rapid.

Considering the overwhelming evidence that GMA is one of the most effective predictors of job performance, why have organizations not adopted GMA testing wholesale? One explanation may be that business professionals, relying on anecdotal data (i.e., stories of individuals who are fired due to “personality conflicts”), conclude that personality factors are more important than cognitive factors. However, research contradicts this, showing that the validity of GMA is 60% to 80% larger than the validity of conscientiousness, which has the most validity evidence of the Big Five Personality traits in predicting performance (Schmidt & Hunter, 2004).

In a 2001 survey of HR Practitioners, Terpstra and Rozell uncovered two dominant reasons that practitioners have not incorporated cognitive ability testing into their selection procedures—beliefs about the usefulness of the practice and legal concerns or fears that the practice would lead to charges of discrimination. Concerns of adverse impact in cognitive testing are valid—certain types of intelligence can be affected by social class and educational advantages (Furnham et al., 2007). It seems counterintuitive, however, that a company would be reluctant to assess GMA (the best predictor of employee performance) for this reason yet continue to conduct interviews (which are less effective) and make hiring decisions based on subjective impressions, which also run a high risk of being subject to bias. In fact, unstructured interviews have been found to have adverse impact against three protected groups—the handicapped, older applicants, and women (Terpstra & Rozell, 1997). Regardless, concerns over adverse impact have

led many professionals to seek alternate types of selection methods, including personality, emotional intelligence, and integrity testing (Scroggins et al., 2008).

Personality Assessments

Most researchers have generally come to accept the Five Factor Model as a universal, all-encompassing model of personality structure (Furnham et al., 2007; Judge et al., 1997), and the stability of this model has been established across applications and geography (Tsaousis & Nikolaou, 2001). The five traits are:

- Emotional Stability—tendency to experience negative emotions such as anxiety, depression, and anger
- Extraversion—tendency toward social behavior, activity, and assertiveness
- Openness to Experience—preference for novel ideas and experiences
- Agreeableness—tendency toward friendly, considerate behavior
- Conscientiousness—tendency to be efficient, determined, responsible, and persistent

Although most researchers accept that scores on personality tests are somehow related to performance (Aronson & Reilly, 2006; Murphy, 2000), the extent to which personality is a valid predictor of performance, especially in a selection context, is not as clear cut (Scroggins, Thomas, & Morris, 2009). The dispersion of opinions among researchers result in vastly different “statements of fact” ranging from “...personality measures do not seem to have much value as predictors of job performance” (Morgeson et al., 2007b, p.1035) to “those who label personality tests in employment selection as having low validities and limited utilities are simply wrong” (Scroggins et al., 2009, p.71). For a highly informative and somewhat “heated” look at the status of personality

test in a selection context, readers are encouraged to refer to the article “Reconsidering the Use of Personality Tests in Personnel Selection Contexts” (Morgeson, Campion, Dipboye, Hollenbeck, Murphy, & Schmitt, 2007a). This article is a recap of a panel discussion that occurred at the 2004 Society for Industrial and Organizational Psychology conference. The article spurred two rebuttal articles--Ones et al. (2007) and Tett & Christiansen (2007)--and a response to the rebuttals by Morgeson et al. (2007b). All four articles can be found in Volume 60 of *Personnel Psychology*.

Studies most commonly find that, of the five traits, conscientiousness is the most generalizable predictor of job performance (Aronson & Reilly, 2006; Furnham et al., 2007; Judge, Higgins, Thorensen, & Barrick, 1999; Schmidt & Hunter, 1998; Viswesvaran, Deller, & Ones, 2007). However, Barrick & Mount's (1991) investigation into the Big Five personality factors as predictors of work outcomes reported conscientiousness (the strongest predictor) had a mean correlation of only .22. Oh, Wang, & Mount (2011) confirmed that the validities of self-report measures of personality tend to be “modest (in the .20s range)” (p. 768). The Department of Labor suggests that criterion-related validity coefficients above .35 are likely to be “very beneficial” (U.S. Department of Labor, 1999), though some researchers feel that “even validities of .20 translate to substantial utility gains” (Ones, et al., 2007, p. 1019).

There are divergent findings in scholarly research concerning whether a particular personality trait is universally desirable or whether its attractiveness is dependent on the context. Most studies have concluded that conscientiousness conclusively predicts job performance across roles (Aronson & Reilly, 2006; Barrick & Mount, 1991). It is theorized that conscientiousness is the strongest of the five personality traits in predicting

job performance because highly conscientious people tend to be very responsible and diligent and work hard to acquire necessary job knowledge (Barrick & Mount, 1991; Schmidt & Hunter, 1998). Barrick & Mount (1991) concluded that “it is difficult to conceive of a job in which the traits associated with the Conscientiousness dimension would not contribute to job success” (p. 22).

Others, however, believe that traits are generally found to be desirable *or not* depending on the context. In other words, some believe that personality constructs are curvilinearly, rather than linearly, related to job performance (Berry & Sackett, 2009; Le et al., 2010; Tsaoisis & Kiklolaou, 2001). Even in the case of conscientiousness, some argue that highly conscientious people may excel in self discipline but “may be considered rigid, inflexible, and compulsive perfectionists” (Le et al., 2010, p. 114) and lack the necessary degree of flexibility to adapt to rapidly changing circumstances. Scholars of this persuasion believe it is necessary to match the traits being measured to the job performance criterion in order for them to be beneficial for prediction in a personnel selection context (Dudley, Orvis, Lebiecki, & Cortina, 2006).

Research on the ability of the four other personality traits (besides conscientiousness) to predict job performance generally follows this contextual or “fit” approach, where a particular personality trait is not inherently “good” or “bad” but could become a strength or weakness depending on the nature of the job (Tsaoisis & Nickolaou, 2001). Barrick & Mount (1993) found that managers who are highly conscientious, highly extraverted, and low on “agreeableness” are successful in roles where they have a great deal of autonomy (and unsuccessful in roles where there is little autonomy). Barrick & Mount (1991) also found that extraversion was a valid predictor of

performance in sales and managerial roles where interpersonal interaction is an important part of the job but is not predictive in other skilled/semi-skilled professional roles that require less interpersonal interaction. Tsaousis & Kikolaou (2001) say the “moderating effects” of things such as job type, degree of autonomy, and amount of social interaction need to be considered. They conclude that personality variables are most predictive of job performance when they are matched with occupational requirements and organizational culture, and take into account variables such as colleagues, supervision, job environment, and reward structure.

In considering future workplace needs and trends, some researchers have asserted that jobs are no longer stable entities and that companies need to be cognizant of the rapidly changing nature of work when considering person-job fit (Anderson et al., 2004). In addition, some researchers are proponents of considering not only person-job fit but also person-team and person-organization fit due to the increased focus on team, rather than individual, performance in organizations (Anderson et al., 2004; Burch & Anderson, 2004). However, few tools have been developed to measure these concepts in a selection context. These constructs are also fraught with difficulties as 1) there could be contradictory traits desired as one analyzes the job function as well as team and organization dynamics, and 2) results are likely to be more difficult to generalize to other contexts due to the more intangible nature of measuring teams and organizations (Anderson et al., 2004).

A prevalent area of concern in the realm of personality tests is faking. Although many tests have a built-in measure of distortion (Tsaousis & Kikolaou, 2001; Wanek, 1999), not all do, and many researchers have investigated the potential effect of faking on

the validity of personality assessments. The conclusions are divergent, with many studies concluding that faking is of great concern while others find that faking has negligible effect on the validity of such tests (Morgeson et al., 2007a). Aronson & Reilly (2006) argue that, of more concern than deliberate faking is a person's tendency to adopt a schema, or a filter that allows selective memory based on what they perceive to be the desired traits for the job, in a high-stakes setting, like applying for a job. When completing a personality assessment as part of applying for a specific job, applicants may adopt a schema that causes them to inaccurately reflect what their behavior actually is, which lowers the ability to make valid inferences. This means that applicants may be subliminally presenting an idealized view of themselves skewed in the direction they believe to be desirable based on the position for which they are applying.

Many studies have sought to find a relationship between personality and intelligence. Some conclude that there is a very small, but significant, overlap between personality and GMA measures (i.e., higher conscientiousness and introversion scores predict higher GMA; Furnham et al., 2007). Others conclude that the constructs are not correlated, and therefore personality can offer incremental validity over GMA of up to 18% (Schmidt & Hunter, 2004). Researchers recommend measuring both personality and intelligence to more fully understand an individual's potential and level of competence (Furnham et al., 2007).

Emotional Intelligence.

As organizations seek selection measures that will help predict future job performance, many have become interested in Emotional Intelligence (EI). The interest in EI began when Goleman released his book on the topic in 1995. Many employers seem to

have latched on to the concept, perhaps in part because they have concerns over using traditional GMA testing due to the potential for adverse impact and are seeking a different measure that would be predictive of employee performance (Van Rooy et al., 2005). Whatever the reason, EI has become a popular and widely used tool for selection and training. As of September, 2008, there were at least 57 EI consulting firms, 90 EI training and assessment organizations, 30 EI certification programs, and five EI “universities” (Jospeh & Newman, 2010).

There are currently two very different models, both purporting to measure EI, operating in the marketplace. One model (Mixed) is closely related to personality traits, while the other (Ability) is more closely aligned with GMA. Furthermore, the existing methods for measuring EI vary and include self-reports, multi-source feedback ratings, and performance-based tests (Christiansen et al., 2010). A discussion of the construct and its merit is included here in the “personality” section because the personality-based model is the one used most often in the context of selection.

The Ability Model claims that EI is a form of intelligence (Van Rooy et al., 2005). More specifically, it is “a type of social intelligence that involves the ability to monitor one’s own and others’ emotions, to discriminate among them, and to use the information to guide one’s thinking and actions” (Mayer & Salovey as cited in Byrne, Dominick, Smither, & Reilly, 2007, p. 342). The primary method of assessing this type of EI involves a scenario-based test in which respondents are asked to select the response that they view as being the most emotionally intelligent. Points are awarded based on how many people in the normative sample selected the same response. Some have raised issues with this consensus-based scoring mechanism because the answers chosen by the

most people are not necessarily the *most emotionally intelligent* responses. Therefore, the tool identifies those who are most like the general population, not necessarily those with the highest EI (Christiansen et al., 2010; Van Rooy et al., 2005).

The Mixed Model is dispositional- or trait-based and has less of a cognitive emphasis. These assessments are a conglomeration of traits, dispositions, motivation, skills, competencies, and abilities (Byrne et al., 2007; Van Rooy et al., 2005). Two popular mixed model measures are Bar-On's EQ-i and the ECI. The EQ-i consists of five general factors: interpersonal EQ, intrapersonal EQ, adaptability, stress management, and general mood. The ECI was created to reflect Goleman's EI model and is a self- and other-report of 18 emotional competencies organized into four clusters: self awareness, self management, social awareness, and social skills. Critics of the mixed model claim there is too much overlap with the Big Five Personality traits (Byrne et al., 2007; Christiansen et al., 2010; Van Rooy et al., 2005) and that test publishers have just repackaged personality measures and are "using a catchy new name to sell worthy, old fashioned personality research and prediction" (Mayer as cited in Byrne et al., 2007, p. 343).

Some studies suggest that these concerns are well founded. Looking at the ECI as an example, the measure is generally unrelated to cognitive ability (.09) but significantly correlated with personality, ranging from .23 with agreeableness and openness to experience to .34 with extraversion (Byrne et al., 2007). Christiansen et al. (2010) also found significant overlap between self-report EI and all four of the Big Five Personality traits (extraversion, agreeableness, emotional stability, and openness; $r = .40-.60$). Mayer, Caruso, and Salovey (2000) echo these concerns when stating "the degree of overlap

between self-report scales of EI and already existing personality scales is a matter of legitimate concern. Given the investment that many people are placing in EI, one would want to ensure there is something new about it” (as cited in Christiansen et al., 2010, p. 89).

Ignoring the fact that EI may indeed be a repackaged version of personality measures, some studies have found the construct has acceptable levels of predictive validity (Van Rooy et al., 2005). Van Rooy & Viswesvaran’s 2004 meta-analysis found that the mean correlation between EI measures (of all types) and work-related outcomes was .24 (Byrne et al., 2007). Byrne et al.’s 2007 study found that ECI scores were positively related to several work-related outcomes, but not after controlling for age and personality. ECI self ratings did explain significant variance (12%) in co-workers’ ratings of managerial skills.

In terms of overall utility, “surprisingly little empirical research exists that demonstrates that EI is related to important organizational outcomes” (Christiansen et al., 2010, p. 87). Christiansen et al.’s (2010) study found that EI explained relatively little variance in job performance that was not already accounted for by other established measures, primarily GMA and conscientiousness. Also, because EI is a composite measure, it does not take into account that some traits may be more or less desirable in certain contexts. Joseph & Newman (2010) “recommend that practitioners use caution when choosing a measure of EI...due to their unknown content and theoretical value” (p. 72) and point out that EI is yet another area where there appears to be a gap between what practitioners believe (as evidenced by the widespread use of EI measures) and what science has been able to support conclusively with data.

Integrity Tests.

Counterproductive workplace behavior (CWB), including theft, absenteeism, and violence, result in significant costs to organizations (Dilchert, Ones, Davis, & Rostow, 2007). The Michigan Employability Survey (Michigan Department of Education, 1989) found that “of 86 employee qualities ranked for importance in entry level employment by over 3,000 employers, 7 of the top 8 qualities were related to integrity, trustworthiness, conscientiousness, and related qualities” (as cited in Ones et al., 1993, p. 697). “Integrity” is a general term that typically encompasses traits such as honesty, dependability, trustworthiness, conscientiousness, and reliability. Some instruments include other traits of interest such as violence tendency, work ethic, hostility, energy level, or substance abuse propensity. The main goal of integrity tests is to measure likelihood of theft or other CWB (Wanek, 1999), though many studies also look at the relationship between scores on integrity tests and overall job performance (Ones et al., 1993).

There was a surge of interest in paper-and-pencil integrity testing after the Employee Polygraph Protection Act of 1988 prohibited employers from using pre-employment polygraphs in most industries. Massachusetts’s and Rhode Island’s Polygraph Acts include language that also restricts or prohibits the use of written examinations to measure honesty or truthfulness. In the rest of the United States, however, there are no legal restrictions on the use of integrity tests, and they have been found not to discriminate or cause adverse impact against any protected group as defined by Title VII of the Civil Rights Act (Sackett et al., 1989).

Integrity tests can be classified into two categories: overt and personality-oriented. Overt tests are also referred to as “clear purpose tests” because the items are

obviously constructed to inquire about counterproductive or undesirable attitudes and behaviors. Overt tests often have two sections: an admissions section and an attitude section (Wanek, 1999). The admissions section directly inquires about past illegal or dishonest behaviors. A sample question might probe into the approximate dollar amount of all the items one has taken from their place of employment without permission in the past year, as an example. The attitude section is designed to measure attitudes and beliefs towards theft and other counterproductive behaviors. Items may ask respondents to answer questions about how they might rationalize theft, what degree of punishment might be appropriate in the case of a particular offense, and how pervasive they believe certain counterproductive behaviors or attitudes to be.

Personality or “veiled-purpose/covert/disguised purpose” tests measure personality constructs such as dependability, social conformity, thrill-seeking, conscientiousness, and issues with authority (Ones et al., 1993; Sackett et al., 1989; Wanek, 1999). They are less transparent in their attempt to measure honesty or theft and are constructed similarly to typical personality assessments, which may be perceived as less offensive by job applicants (Sackett et al., 1989). Although the definitions of overt and personality-oriented tests seem clear-cut, in practice tests may have elements of both approaches (Wanek, 1999). Integrity tests are also delineated by whether they are designed to measure narrow criteria (i.e., theft) or broad criteria (i.e., general disruptive or rule-breaking behavior like disciplinary problems, excessive tardiness or absence, violence, substance abuse, property damage; Ones et al., 1993). Counterproductive behaviors can be measured through “external criteria” (organization records of disciplinary actions,

attendance records, theft, etc.) or “self-report criteria” (admission of past illegal or unethical behavior) (Ones et al., 1993).

Historically, the validity of integrity tests has been assessed through means such as correlating scores with polygraph tests, comparing scores among groups of people thought to differ substantially on the construct (such as convicted felons vs. monks or nuns), or tracking the change in outcomes, such as inventory shrinkage, over time after implementing an integrity test. More recently, researchers have measured integrity tests’ ability to predict broader outcomes such as absenteeism, disciplinary actions, grievances, and terminations (Hogan & Hogan, 1989; Sackett et al., 1989; Wanek, 1999).

Several studies have suggested that professionally developed integrity tests are valid predictors of counterproductive behaviors (Ones et al., 1993; Wanek, 1999). According to Ones et al.’s Meta-Analysis, overt integrity tests’ validity in predicting overall job performance was .33 (.35 for personality-based integrity tests). The mean validity of overt tests in predicting disruptive behaviors was .55, (.32 for personality-based integrity tests). The average expected validity of integrity tests in predicting supervisory ratings of overall job performance is .41 (Ones et al., 1993; Schmidt & Hunter, 1998; Schmidt & Hunter, 2004). There has been some discussion around interpretational issues due to the fact that companies implementing integrity testing rarely utilize a control group. Therefore, there is some uncertainty whether the screening ability of the integrity test is solely responsible for improved outcomes or if the company’s implementation of this additional screening method sends a message about the company’s tolerance level of theft and other counterproductive behaviors, which becomes a confounding variable (Sackett et al., 1989). From a practical standpoint,

however, one could conclude that the objectives the organization set out to achieve by implementing an integrity test are achieved regardless of the underlying cause.

Several studies have investigated the relationship between the construct of integrity and the Big Five Personality traits. Judge et al. (1997) found that Extraversion and Conscientiousness were moderately strong predictors of employee absence (one of the outcomes purportedly measured by many integrity tests). Bernardin (1977) used the 16PF personality assessment and found that conscientiousness and anxiety account for a small but significant amount of variance in organizational withdrawal behaviors such as turnover and absenteeism. A significant overlap has also been found between integrity and conscientiousness, which encompasses traits such as dependability, carefulness, and responsibility (Ones et al., 1993). However, overall validities are higher for integrity assessments than for assessments of conscientiousness (Schmidt & Hunter, 2004).

Numerous studies have looked at the correlation between integrity measures and measures of GMA (Werner, Jones, & Steffy, 1989). Because the relationship between integrity and GMA in such studies was found to be zero, the expected incremental validity that can be gained from adding measures of integrity to GMA tests can be calculated. Most studies show an approximate incremental validity of 27% in validity, over using cognitive measures alone (Schmidt & Hunter, 1998; Schmidt & Hunter, 2004; Wanek, 1999).

Summary of Literature Review

In reviewing the body of scholarly literature surrounding assessments of GMA and personality, research suggests that GMA is one of the strongest predictors of job performance (Christiansen et al., 2010; Furnham et al., 2007; Schmidt & Hunter, 1998;

Schmidt & Hunter, 2004; Terpstra & Rozell, 1993; Wanek, 1999), but the presence of adverse impact in using this construct has caused many practitioners to look to alternate or additional constructs for use in employee selection (Furnham et al., 2007). Concerning personality tests, conscientiousness is generally found to be the strongest predictor of performance of the Big Five personality traits (Aronson & Reilly, 2006; Furnham et al., 2007; Judge et al., 1999; Schmidt & Hunter, 1998; Viswesvaran et al., 2007), and numerous studies have shown various personality traits, or combinations of traits, to be effective predictors in certain types of jobs (Barrick & Mount, 1991; Barrick & Mount, 1993; Tsaousis & Kikolaou, 2001). Research also suggests that integrity tests are effective predictors of both counterproductive workplace behaviors and overall job performance (Ones et al., 1993; Wanek, 1999).

The above findings suggest that employers would benefit from relying more heavily on selection methods with proven predictive validity. According to Schmidt & Hunter (1998), one strategy is to assess GMA in combination with integrity, which would achieve 65% of the “maximum possible practical value (utility)” (Schmidt & Hunter, 1998, p. 267). Employers should also consider the role that personality testing may play in their selection processes and might consider doing some analysis to determine whether or not certain personality traits lend themselves to better job performance in certain roles within their organizations. When taking into consideration the costs associated with hiring personnel, and with the cost of turnover when poor selection decisions are made, utilizing more valid selection methods should yield a substantial return-on-investment (utility). Strategic human resource professionals can have a significant effect on business outcomes by creating selection systems with high predictive validity.

In light of the general findings of scholarly research relative to GMA and personality tests, this study seeks to better understand what information practitioners receive relative to selection assessments via practitioner literature as a way of exploring possible reasons for the science-practitioner gap in the area of selection. Both scholarly and practitioner articles will be analyzed to determine the extent to which they are aligned in their coverage of workplace assessments relative to five research questions or “themes.” The five questions, as well as what conclusions may be drawn as a result of the analysis, are as follows:

1. *What is the main purpose of the article?* Are practitioner and scholarly publications focused on the same issues relative to selection assessments?
2. *What category of workplace selection assessment is discussed?* Are HR practitioners receiving information about the same types of constructs and measures that scholars are researching?
3. *Are any commercially available assessment instruments mentioned specifically? If so, which ones?* Are scholars and practitioners using and/or being exposed to the same instruments?
4. *How is validity discussed?* Analyzing how validity is discussed and reported in scholarly versus practitioner literature will allow the extent to which both groups are “speaking the same language” to be determined and, to the extent that they are not speaking the same language, may provide insight into a reason for the reported “knowledge gap” relative to selection assessments.
5. *To what extent and in what manner does the article address the concept of utility (i.e. the expected benefit from using assessments)?* To what extent do

scholars and practitioners share similar rationales and justifications for using assessment instruments in a selection context?

CHAPTER 3

METHODOLOGY

The purpose of the present study was to compare the messages contained in practitioner literature and scholarly literature concerning self-report employee selection assessment instruments. To determine the extent to which scholarly and practitioner publications were aligned in their coverage of workplace assessments, a qualitative content analysis (Zhang & Wildemuth, 2009) of articles pertaining to selection assessments found in two practitioner-oriented publications and two of the top HR-oriented scholarly journals over a five-year period was conducted. Qualitative content analysis has been described as “any data reduction and sense-making effort that takes a volume of qualitative material and attempts to identify core consistencies and meanings” (Zhang & Wildemuth, 2009, p. 1). Articles were analyzed for content related to five research questions, or “themes”—purpose of the article, type of assessment instrument discussed, specific instruments mentioned, discussion of validity, and discussion of utility. Details related to the rationale for what publications were selected for inclusion in the analysis and for what timeframe follows. The criteria used for deciding whether a particular article should be included or excluded from analysis is also discussed, along with the procedure used for the content analysis.

Scholar and Practitioner Journal Selection

The Journal of Applied Psychology and *Personnel Psychology* have been identified as the two most prestigious, “flagship” (Judge, Lepine, & Rich, 2006; Zickar &

Highhouse, 2001), and HR-generalist-relevant scholarly journals. These two journals are considered “primary outlets” for research on personality tests in a selection context (Morgeson et al., 2007a, p.687), and their selection for review is consistent with similar studies (e.g., Deadrick & Gibson, 2007; Rynes et al., 2007). Other scholarly publications were considered for inclusion but subsequently rejected. For example, the *Academy of Management Journal*, while considered very prestigious (Zickar & Highhouse, 2001), is not especially geared toward Human Resource professionals, and a preliminary key word search revealed that no articles relevant to this study were included in this journal in the most recent four years of publication. Also, the selection of HR *generalist* oriented publications enabled the relative emphasis on selection assessments within the broad range of topics of interest to HR scholars and practitioners to be analyzed.

The two practitioner-oriented publications that were selected for review were *HR Magazine* and *HR Executive*. *HR Magazine* is widely recognized as the premier HR practitioner journal with a circulation rate of over 250,000 and has been used for this reason in similar previous studies (Deadrick & Gibson, 2007; Rynes et al., 2007). Rynes et al. (2002) found that *HR Magazine* was the only HR publication that respondent reported reading “more than sometimes.” The second magazine, *HR Executive*, was selected because it has the next highest circulation rate (75,034) of HR generalist practitioner publications.

Timeframe

The selection of the time-frame for review was primarily motivated by a desire to look at recent publications and assess the *current* dialog occurring in these publications relative to selection assessments. The years under review (2006-2011) have been

primarily dominated by economic recession, which creates unique challenges for employers in the area of employee selection, retention, and talent management. Because the goal of this study is not to look at trends over time, a lengthy time span was not deemed critical. Also, because this study refines, builds upon, and continues two aforementioned studies by Deadrick & Gibson (2007; pertaining to the gap in interests between practitioners and scholars as revealed in the topics of articles published in practitioner and scholarly journals) and Rynes et al. (2007; pertaining to the knowledge gap between scholarly findings and practitioner beliefs), it is appropriate to begin analysis where they left off, which for both studies was in 2005. As a result, this study includes content published from January, 2006, through September, 2011.

Procedure

Online archives for each publication were accessed, and article titles and abstracts were systematically reviewed, issue by issue, to identify relevant articles. Any title or abstract that alluded to hiring, selection, recruitment, testing, personality, GMA, integrity, etc. were further reviewed to determine relevance. Any article that addressed *paper-and-pencil (or internet based) self-report tests/inventories/instruments* used for employee selection was included. A few articles that were initially selected for inclusion were subsequently eliminated when further review revealed that the article was exclusively concerned with something like refining a construct or statistical method (e.g., the Five Factor Model of personality, methods for conducting meta-analyses) rather than addressing the use of an assessment instrument as a predictor of job performance in employee selection.

In reviewing the content of each publication, the content type of interest was “articles.” Articles are presumably written to educate or share information and are based on some piece of knowledge or research. For this reason, letters to the editor, book reviews, pieces clearly marked as “advertising,” calls for papers, etc. were omitted from the “total article” count. These decisions are consistent with other similar studies such as Deadrick & Gibson (2007) and Rynes et al. (2007).

The final population used in the subsequent analysis was 4,408 articles (771 scholarly, 3,637 practitioner), out of which 49 were deemed relevant to the current study and included in the qualitative content analysis. Of these, 36 articles were from scholarly journals (22 from *Journal of Applied Psychology*, 14 from *Personnel Psychology*), and 13 were from practitioner publications (6 from *HR Magazine*, 7 from *HR Executive*). Table 3 presents the number and relative percentages of articles included from each source. It should be noted that *HR Executive* is published approximately 16 times per year and includes an average of 26.5 articles per issue, and *HR Magazine* is published monthly and includes an average of 18.5 articles per issue. In contrast, *Journal of Applied Psychology* is published only six times per year and includes an average of 18.5 articles per issue, and *Personnel Psychology* is published only quarterly and includes an average of 8.2 “academic journal” articles per issue. This publication schedule affects the quantity of scholarly versus practitioner articles included in the review. Also, the reader should note that there were more than 49 articles in these publications that discussed *personnel selection*—however, such articles discussed methods other than paper-and-pencil self-report instruments, like interviews, assessment centers, etc. Articles included in the content analysis are marked with an asterisk in the reference section after Chapter 5 of

this thesis. The Appendix also lists the content relevant to each research question for each article.

Table 3

Number of Articles Included From Each Source

	HR Exec	HR Mag	<i>Pract.</i> <i>Total</i>	PP	JAP	<i>Schol.</i> <i>Total</i>	Totals
Total Articles in Timeframe	2,361	1,276	3,637	182	589	771	4,408
Total Relevant/ Included	7	6	13	14	22	36	49
% of total articles included	0.30%	0.47%	0.35%	7.69%	3.74%	4.67%	1.11%

Note. HRExec= *HR Executive*; HRMag=*HR Magazine*; Pract. Total=Total from Practitioner Publications; PP=*Personnel Psychology*; JAP=*Journal of Applied Psychology*; Schol. Total=Total from Scholarly Publications

Content Analysis

Qualitative content analysis has been described as “any data reduction and sense-making effort that takes a volume of qualitative material and attempts to identify core consistencies and meanings” (Zhang & Wildemuth, 2009, p. 1). This method enables a researcher to identify, analyze, and report patterns or “themes” within data (Braun & Clarke, 2006; Morreale, Backlund, Hay, & Moore, 2011). Because this study is building on previous research and seeks to answer specific questions about the coverage of workplace assessments, the analysis began with a deductive (rather than inductive) approach, which is appropriate when the analysis is to be “driven by the researcher’s theoretical or analytic interest in the area” (Braun & Clark, 2006, p.84).

A worksheet consisting of the five research questions or “themes” to be analyzed was completed for each article. The resulting content was then analyzed, doing further

coding where necessary. Whereas two themes (theme #2—type of assessment mentioned, and theme #3—name of particular instruments) merely required tallying for analysis, the remaining three themes (relative to “article purpose,” “validity,” and “utility”) were subjected to a second cycle inductive coding process—where “themes and categories emerge from the data,” (Zhang & Wildemuth, 2009, p.2)—in order to further classify the content into comprehensive and mutually exclusive categories that facilitated the comparison of findings between sources. The ultimate codes and categories are further explained in the Results section.

CHAPTER 4

RESULTS

To better understand potential causes of the scholar-practitioner gap relative to selection assessments, qualitative content analysis was used to analyze 49 articles contained in two scholarly and two practitioner publications in the last five years. Articles were analyzed relative to five research questions, or “themes”—purpose of the article, category of assessment discussed, specific instruments mentioned, discussion of validity, and discussion of utility. The results of the analysis relative to each theme are presented below.

Theme 1: Purpose of the Article

The first research question (or theme) concerns the main purpose of the article. Coding each article as to its primary message or purpose facilitates the examination of the degree of consistency between practitioner and scholarly publications in their focus as it relates to selection assessments. The main purpose of each article, which was usually evident in either the title or first few sentences of the article or abstract, was succinctly summarized. An inductive coding strategy was then used to group articles with similar purposes together and a label was assigned to each group. An article whose main purpose was to address the issue of faking on tests was labeled as “Faking.” Any article that was primarily discussing the issue of bias or adverse impact as it relates to assessments was labeled “Bias/Adverse Impact.” Similarly, articles that explored some construct’s ability to predict counterproductive workplace behaviors (CWB) or job performance were

grouped together, as were those providing an overview or review of research on a particular category of test.

For the remainder of the groups it may not be as obvious to the reader what kinds of articles comprise the group; therefore, an explanation is provided here. The group “How to Screen for (x)” was comprised of practitioner articles that explored various ways to screen for sales aptitude, team engagement raising characteristics, CWB, or other interpersonal skills—in other words, they were working *backwards* from some desired end result and looking at a variety of ways to screen for it, as opposed to the articles included in the “Predicting CWBs/Job Performance” categories which *start* with a construct of interest and explore the statistical correlation between that construct and the stated outcome of either CWBs or Job Performance. Two articles that discussed how factors such as technology and globalization have affected the assessment industry were labeled “The State of Assessments.” Articles that advocated the use of assessments due to economic drivers, the need for data-driven decision making, or based on the potential for return-on-investment were labeled “Why Use Assessments.” The group “Enhancing Measures” was comprised of articles with purposes such as reviewing range restriction processes for validity estimates, proposing a new model of a construct (i.e., Emotional Intelligence or the Five Factor Model of Personality), or exploring the use of new response scales. Two articles, which concerned the effects of retesting and the presence of “publication bias” in materials from test publishers, were found not to fit within any of these categories and were therefore assigned to a “miscellaneous” category. The number of articles assigned to each category from each source are presented in Table 4.

Table 4

Purpose of Articles by Source

	HRExec	HRMag	Pract. Total	PP	JAP	Schol. Total
Why Use Assessments?	2 (29%)	1 (17%)	3 (23%)			
The State of Assessments	1 (14%)	1 (17%)	2 (15%)			
Bias/Adverse Impact		1 (17%)	1 (8%)	1 (8%)	2 (9%)	3 (8%)
How to Screen for (x)	1 (14%)	2 (33%)	3 (23%)			
Overview—Integrity		1 (17%)	1 (8%)	1 (8%)		1 (3%)
Overview—Personality				4 (28%)		4 (11%)
Predicting CWBs	2 (29%)		2 (15%)	2 (14%)	3 (14%)	5 (14%)
Predicting Job Performance				2 (14%)	5 (23%)	7 (19%)
Faking	1 (14%)		1 (8%)		6 (27%)	6 (17%)
Enhancing Measures				2 (14%)	6 (27%)	8 (22%)
Miscellaneous				2 (14%)		2 (6%)

Note. HRExec= *HR Executive*; HRMag=*HR Magazine*; Pract. Total=Total from Practitioner Publications; PP=*Personnel Psychology*; JAP=*Journal of Applied Psychology*; Schol. Total=Total from Scholarly Publications

As is shown in Table 4, although there were areas of overlap, there was also a significant amount of segmentation in the purposes of scholarly versus practitioner articles. For example, only practitioner publications had articles that spoke to the state of the assessment industry as a whole, advocated the use of assessments in general, or discussed various tools and techniques for screening for some desired trait. Conversely,

articles that examined the statistical correlations between some construct and some outcome or articles that scientifically examined ways to enhance measures (i.e., statistical processes, response scales, or measurement methods) were limited to scholarly publications.

Theme 2: Category of Assessment Discussed

By categorizing and counting the references to various types of selection assessments (e.g., intelligence/aptitude, personality, honesty/integrity) it is possible to determine whether HR practitioners are receiving information about the same types of constructs and measures that scholars are researching. For example, Rynes and colleagues (2007) found no articles concerning GMA or intelligence testing in *HR Magazine* during the timeframe of their study (2000-2005) despite the fact that GMA is a widely studied construct in academic circles. Therefore, in the current study, the number of assessment categories referenced in each source was recorded and is presented in Table 5.

Table 5

Category of Assessment Referenced by Source

	HRExec	HRMag	Practitioner Total	PP	JAP	Scholarly Total
GMA	3 (23%)	3 (23%)	6 (46%)	3 (8%)	10 (28%)	13 (36%)
Personality	3 (23%)	3 (23%)	6 (46%)	13 (36%)	17 (47%)	30 (83%)
Integrity	3 (23%)	1 (8%)	4 (31%)	3 (8%)	0 (0%)	3 (8%)
Other	3 (23%)	3 (23%)	6 (46%)	0 (0%)	2 (6%)	2 (6%)

Note. HRExec= *HR Executive*; HRMag=*HR Magazine*; PP=*Personnel Psychology*; JAP=*Journal of Applied Psychology*. Percentages do not add up to 100% because articles could be assigned to more than one category.

As is shown in Table 5, practitioner publications reviewed in this study were fairly even in their coverage of various types of assessment instruments with six articles referencing personality and GMA-type tests, and four referencing integrity tests. Six articles also referenced other types of assessments (e.g., skills, values, or knowledge/certification types of assessments). Scholarly publications showed a distinctly heavier emphasis on personality measures with 30 out of 36 articles addressing personality tests. GMA received moderate attention (13 out of 36 articles), and integrity received very little attention with only 3 articles discussing these types of tests (although several of the personality articles were discussing the ability of personality traits to predict CWB).

Theme 3: Specific Instruments Mentioned

The research question pertaining to the third theme investigated pertained to the extent to which articles from scholarly versus practitioner publications assessed specifically-named selection instruments. Although a few commercially available instruments were mentioned by name in practitioner literature, more often test *publishers* were referenced. This happened in two ways: 1) the article would allude to tools or instruments available from a particular publisher (i.e., a dependability measure from SHL), or 2) the article would quote an executive from a test publisher (implying their expertise). The publishers referenced more than once (and the number of times they were referenced in separate articles) in practitioner literature were: Previsor (7), Kenexa (4), DDI (4), SHL (2), Select International (2), Pearson (2), and PDI (2). The instruments referenced more than once in scholarly articles (and the number of times they were referenced in separate articles) were: NEO Personality Inventory (9), California

Personality Inventory (8), Hogan Personality Inventory (7), 16PF (4), MMPI (3), Wonderlic (2), Jackson Personality Inventory (2), and the Armed Forces Qualification Test (2). There was found to be no overlap in the specific instruments referenced consistently in scholarly versus practitioner articles in this study. Hogan is the only test publisher that is referenced in scholarly literature that also appeared in practitioner literature, although it only appeared in practitioner literature once

Theme 4: Validity

The focus of the fourth research question was to evaluate the use of the term “validity” in the articles pertaining to selection assessments in the scholarly versus the practitioner journal articles. In order for practitioners to select a pre-employment assessment instrument, they must possess some knowledge regarding the existence of and importance of validity, as well as how the validity of an instrument is determined. Analyzing how validity is discussed and reported in scholarly versus practitioner literature allows evaluation of the extent to which both groups are “speaking the same language” and, to the extent that they are not speaking the same language, may provide insight into a reason for the reported “knowledge gap” relative to selection assessments.

To this end, any mention of the word validity (or any derivation of this word) was documented, along with the context in which the reference was made. The resulting content was then analyzed and grouped into six categories: “none,” “implied,” “isolated word use,” “validation study alluded to,” “validity discussed,” and “validity reported.” These categories are further defined below.

- *None*—these articles included no use of any form of the word “validity” and no discussion surrounding accuracy, probability, predictiveness, or other concepts that could be related to validity.
- *Implied*—these articles included no use of any form of the word “validity.” However, phrases such as ‘detects certain traits with a remarkable degree of *accuracy*,’ ‘helps increase the *probability* of making correct decisions,’ or ‘testing is one of the most *predictive* parts of the application process’ were used, which hints at the concept of validity.
- *Isolated Word Use*—these articles included the word “validity” but lacked any explanation or context to help the reader infer meaning. For example: ‘as long as tests have been validated, they are acceptable for use,’ ‘[company] is a provider of validated employee screening products,’ or ‘properly validated assessments have long been viewed as an important arrow in the performance management quiver.’
- *Validation Study Alluded To*—without in-depth reporting of study results, these articles referred to a “validation study” that showed that higher test scores correlated with some desirable behavior like tenure, sales, or job performance.
- *Validity Discussed*—in these articles, some explanation, definition, or discussion accompanied the use of the word “validity.” For example: ‘test publishers conduct research to demonstrate the validity of tests—that is the accuracy with which the test predicts future job performance,’ or ‘the strongest validation evidence is obtained by demonstrating that people who score higher on tests actually perform better on the job.’ Articles were also included in this category if they included more advanced content or language surrounding validity (i.e.: ‘a sample of over

300 is needed for statistical validity’ or ‘each competency is weighted according to its predictive value’).

- *Validity Reported*—these articles actually reported detailed results of validation studies and/or included some numeric reporting of validity (i.e., coefficients).

These six categories can be arranged on a continuum of “sophistication” (see Figure 1) with the far left side representing a very unsophisticated treatment of validity, the middle representing a moderate or “layman’s” treatment of validity, and the right representing a fairly to very sophisticated, or “advanced,” treatment of validity. Articles were coded based on the highest level of sophistication obtained, meaning that each article was assigned to only one category.

Figure 1

Continuum of Sophistication for Validity Discussion

The number and relative percentage of articles falling into each of these six categories by source is presented in Table 6. As can be seen in Table 6, there was found to be no overlap between scholarly and practitioner articles in the level of sophistication used in their discussion of validity.

Table 6

Number of Articles Per Validity Category by Source

	HRExec	HRMag	Pract. Total	PP	JAP	Schol. Total
None	3 (43%)		3 (23%)			
Implied	1 (14.25%)		1 (8%)			
Isolated Word Use	1 (14.25%)		1 (8%)			
Validation Study eluded to	1 (14.25%)	2 (33%)	3 (23%)			
Validity Discussed	1 (14.25%)	4 (67%)	5 (38%)			
Validity Reported				14 (100%)	22 (100%)	36 (100%)

Note. HRExec= *HR Executive*; HRMag=*HR Magazine*; Pract. Total=Total from Practitioner Publications; PP=*Personnel Psychology*; JAP=*Journal of Applied Psychology*; Schol. Total=Total from Scholarly Publications

Theme 5: Utility

The focus of the fifth research question was on the theme of “utility” and how this was addressed in the scholarly and practitioner-oriented journal articles on selection assessments. How utility, or the “practical economic value” (Schmidt & Hunter, 1998, p. 262), of using selection assessments is discussed in practitioner and scholarly sources sheds light on the extent to which scholars and practitioners share similar rationales and justifications for using assessment instruments in a selection context.

The process for analyzing utility was similar to that used for validity. Each reference to utility (defined here as some expected benefit from using assessments) was recorded then subsequently grouped together with similar statements and broken into

categories. This topic also lent itself to categorization along a continuum of sophistication ranging from “none” to “case study.”

Practitioner and scholarly articles report on utility very differently—scholarly articles regularly report validity coefficients (r values) that reflect a construct’s ability to predict some outcome (i.e., job performance, counterproductive workplace behaviors). Because “how validity is reported” is a separate question in this study, this theme seeks to categorize articles on how they speak to utility (expected benefit from using a selection instrument) independent of merely reporting validity and assuming that is sufficient evidence from which utility can be surmised. Therefore, in this category, the interest is in some explicit statement about the benefit of using an assessment instrument. The categories established from the data are detailed below, and the number of articles falling into each category from each source are presented in Table 7.

- *None:* These articles made no reference to any expected benefit from using assessments for employee selection. In the case of scholarly articles, most of which report some sort of validity coefficient, articles were categorized as “none” if they made no explicit statement regarding the benefit of using assessments (in most cases these articles were focused exclusively on studying bias, retest effects, faking, etc., rather than a construct’s ability to predict some criterion).
- *Implied:* The text in these articles implied a vague benefit of using assessments without relating it to any specific desired outcome. For example, ‘good assessments are the most effective and efficient tools HR has for making hiring decisions,’ ‘assessments add tremendous value,’ or ‘assessments are useful when a company has a big applicant pool and needs to find the best fitting candidate.’

- *Assumed:* These articles stated that assessments are effective in achieving some desired outcome, without actually providing any supporting evidence (other than r values). For example: ‘assessments have been repeatedly shown to yield important bottom line outcomes such as increased productivity, reduced turnover, etc.,’ ‘benefits include retention and higher profits, reduced turnover, increased productivity,’ or ‘whether the goal is improving the cost-per-hire or the quality-of-hire, assessments are an indispensable piece of the decision making puzzle.’ In the case of scholarly articles, because most included articles are reporting a construct’s ability to predict some desired outcome (i.e., job performance), many automatically fell into this category. In other words, it is “assumed” that the reader understands that, by virtue of reporting an r value, there is some utility in using the construct, since no other explicit statement is made regarding the expected benefit.
- *Case Study:* These articles referenced or described a case study where the implementation of an assessment process resulted in some documented outcome. For example, ‘those with better assessment scores signed 11 times more clients per month,’ ‘turnover was reduced by two-thirds,’ or ‘high scores on conscientiousness correlated with \$50 thousand less stock shrinkage, which adds up to \$78 million in company-wide savings.’ In other words, it included an explicit reference to the *impact* of using the tool (beyond just reporting the r value).

Table 7

Number of Articles Per Utility Category by Source

	HRExec	HRMag	Pract. Total	PP	JAP	Schol. Total
None	2 (29%)		2 (15%)	2 (14%)	5 (23%)	7 (19%)
Implied	3 (42%)		3 (23%)		1 (4%)	1 (3%)
Assumed		1 (17%)	1 (8%)	12 (86%)	16 (73%)	28 (78%)
Case Study	2 (29%)	5 (83%)	7 (54%)			

Note. HRExec= *HR Executive*; HRMag=*HR Magazine*; Pract. Total=Total from Practitioner Publications; PP=*Personnel Psychology*; JAP=*Journal of Applied Psychology*; Schol. Total=Total from Scholarly Publications

Perhaps the most notable finding relative to this theme is that none of the scholarly articles included in this review were assigned to the “Case Study” category, meaning that there was no discussion in scholarly sources surrounding specific bottom-line outcomes achieved by using a selection assessment. In contrast, the majority (7 out of 13) of practitioner articles took this approach to discussing utility.

Summary

By conducting qualitative content analysis on 36 scholarly articles and 13 practitioner articles pertaining to selection assessments, the degree of consistency with which publications from each orientation discuss selection assessments could be analyzed. It was found that there were distinctly different focuses in selection-assessment-related articles in scholarly versus practitioner publications (theme 1). It was also found that scholarly publications showed a distinctly heavier focus on personality assessments than GMA or integrity, while practitioner articles were fairly even in their coverage of various categories of assessments (theme 2). There was found to be no

overlap in specific instruments mentioned consistently in scholarly versus practitioner publications (theme 3). It was also found that scholarly and practitioner articles take significantly different approaches when discussing validity (theme 4) and utility (theme 5). A discussion of these results follows.

CHAPTER 5

DISCUSSION

Qualitative content analysis of 36 articles from two scholarly publications and 13 articles from two practitioner publications in the last six years concerning self-report selection assessment instruments revealed notable differences in terms of the overall purpose of the articles, category of assessments discussed, specific instruments mentioned, discussions of validity, and discussions of utility. The findings relative to each of these five themes, along with implications for scholars and practitioners, limitations of the current study, and suggestions for future research are discussed in the following sections.

Theme 1: Purpose of the Article

A different focus in practitioner versus scholarly publications was found in this study. Practitioner articles tended to have a *broad*, “big picture” focus (i.e., Why should a company consider using assessments? What is going on in the assessment industry?). This may reflect the desire of practitioners to read articles that are more global in application and more focused on the concerns of the end users (i.e., the benefits or potential return-on-investment of using assessments or the effectiveness of such tools in screening for some desired trait).

On the other hand, while scholarly publications also included several broad “overview” articles that addressed the current state of research related to selection assessments, a majority of the articles from the scholarly publications were more *specific*

(rather than broad) in their focus. The areas of heaviest emphasis in scholarly publications were how specific constructs correlated with or predicted counterproductive workplace behaviors (CWB) or job performance, the impact of faking on personality assessments, and articles discussing how to enhance specific measures (i.e., adding “context” to personality assessments, exploring alternate response scales, etc.). This is consistent with the scholarly goals of theory-building and construct refinement.

These results echo Deadrick & Gibson’s (2007) finding that HR academics and HR professionals value different types of knowledge, as is evidenced by the different focus of practitioner versus scholarly HR publications. Although Deadrick & Gibson’s (2007) analysis was broader in nature, they concluded that HR professionals tend to be more interested in topics that are relevant to the “practical nature of the day-to-day job demands” whereas HR scholars are primarily interested in “creating generalizable research” (p. 136). This “interest gap” is likely to persist as long as each group is driven and incentivized to develop and acquire knowledge in each of these separate areas.

Theme 2: Category of Assessment Discussed

It is difficult to draw any definitive conclusions about the quantity and dispersion of discussion surrounding different categories or types of assessments between the two types of journals investigated in this study. A preponderance of the scholarly articles that fit the scope of this study were found to be about personality tests, with less than half as many articles discussing general mental ability (GMA). It is noteworthy to compare these findings to what Rynes et al. (2007) found when analyzing the same two scholarly journals for articles relevant to GMA and personality for selection from 2000-2005. In their study, *Journal of Applied Psychology* and *Personnel Psychology* devoted 3.2% and

6.3% respectively of their articles to GMA and 5.9% and 6.3% respectively to personality. This means that during the timeframe of Rynes et al.'s (2007) study personality was also given a majority of the attention but GMA was given much more attention than in the current study relative to personality. It is established that scholars place a high degree of importance on the role of GMA in predicting performance (Rynes et al., 2007), so it is somewhat confounding why the relative number of scholarly articles pertaining to GMA in selection appears to have decreased in the past 6 years.

As for practitioner publications, the Rynes et al. (2007) study found no articles in *HR Magazine* pertaining to GMA in the timeframe studied and only three articles (0.4%) regarding the role of personality in selection. In contrast, the current study found the exact same number of articles regarding personality in *HR Magazine* (3), but also found three articles that spoke to the role of GMA in selection, which is a substantial increase. Drawing any conclusions as to why this increase occurred would be pure speculation, but it would be interesting to know (if it were possible to know) if the Rynes et al. (2007) study perhaps had some effect on practitioner publication practices.

Theme 3: Specific Instruments Mentioned

Virtually no overlap was found between scholarly and practitioner publications in the specific assessment instruments mentioned. Hogan is the only test publisher that is referenced in scholarly literature that also appeared in practitioner literature, although it only appeared in practitioner literature once. There are two prominent factors most likely contributing to this situation. First, the vast majority of research conducted on selection instruments is conducted by the test publishers themselves, and that research is often not reported in peer-reviewed journals. Second, because instrument questions and scoring

mechanisms are proprietary, it is often difficult for instruments widely used in the public sector to be analyzed and evaluated by independent researchers, and cooperation from test publishers in this regard is not always easy to obtain (Sackett et al., 1989).

This situation is concerning as it relates to the ability of practitioners to become educated consumers of selection tests. If test publishers are the only entities privy to the exact psychometric properties of an instrument and they are the ones producing all literature relative to the reliability and validity evidence of an instrument, then the extent to which negative results may be suppressed is unknown (Sackett et al., 1989). Studies have confirmed that publication bias exists (McDaniel et al., 2006) and that some test publishers may use sales tactics that are deceptive to some degree (Sackett et al., 1989). Therefore, practitioners really have very little choice when attempting to assess the usefulness of a test, especially if they are not experts in statistical processes, other than to trust the publisher.

In his list of considerations for selecting an assessment instrument, Wanek's (1999) first suggestions are to ensure proven reliability, validity, and absence of adverse impact. Also, the *HR Magazine* article "Effective Assessment" (Zielinski, 2011) includes a list of questions that HR practitioners should ask vendors: *When were validity tests done? For what jobs was the instrument validated? What were the results of the validity test? What was the test coefficient? What was the sample size? Was the test shown to be reliable as well as valid?* (Assessment Trends section, para. 12). Although this is sound advice, it still does not overcome the obstacles of many practitioners not being equipped to contextualize, make sense of, and apply this information, or the fact that the answers to these questions from test publishers may not be entirely accurate.

Theme 4: Validity

In looking at the way that scholarly and practitioner articles discuss validity, there was found to be no overlap between the two sources in terms of the degree of sophistication used in addressing this vital concept. Regarding practitioner publications, no articles reported validity numerically. This is not particularly surprising given most practitioners are not trained to understand statistics. It is also not surprising that all scholarly articles included very advanced and technically complex discussions surrounding validity given that is the expectation of journal editors and the consumers of this literature (other scholars).

Although it is not *surprising* that the two sources differ in their treatment of validity, this does not mean it is not *problematic*. Validity should be among the most important considerations for a HR practitioner when choosing to use a selection assessment (Schmidt & Hunter, 1998; Scroggins, et al., 2008). However, there is evidence to suggest that the predictive validity of a tool plays a relatively small role in practitioner decisions to adopt an assessment process (König et al., 2010). This could possibly be because practitioners do not feel equipped to discuss or interpret validity information when considering implementing a selection instrument and practitioner literature appears to be doing little to address this concern. It is difficult to see how the scholar-practitioner gap can be narrowed when the discussion lacks a common language.

It should be noted that there were two practitioner articles with *relatively* advanced discussions surrounding validity. The first, “Screening: Testing the Limits” (Geisinger, 2008), was written by a Ph.D. who is also the director of the Buros Center on Testing and the editor of *Applied Measurements in Education* among other scholarly

journals. The second, “Your Cheating Heart” (Roberts, 2011b), included statements such as “integrity tests have been validated repeatedly by developers, independent researchers, and HR professionals,” “[integrity tests have been shown to be] statistically valid and predict what they purport to,” and “the higher the validity of the test, the lower the error rate.” This article relied extensively on quotes from Deniz Ones, a highly recognized scholar and researcher in the area of integrity testing.

It would be interesting to know if this apparent attempt to “bridge the gap” in the literature was instigated by the two aforementioned scholars or by the practitioner journals, and whether this effort was in reaction to studies like Deadrick & Gibson (2007) and Rynes et al. (2007) that have challenged both scholars and practitioners to take this kind of initiative. These two articles could potentially be indicative of the beginning of a positive trend toward integrating scholarly perspectives into practitioner literature.

Theme 5: Utility

The differing approaches of scholar versus practitioner publications when it comes to discussing utility are another area in which a gap was evident. Whereas many practitioner articles made some statement about the expected benefit of using selection tests, up to and including fairly detailed case studies where the impact of implementing an instrument was dollarized, none of the scholarly articles included in this review attempted to make any explicit statement as to the potential impact on an organization if it were to use a predictive construct to measure some criterion. The highest “sophistication” level attained in scholarly articles is to report an *r* value and *assume* the reader can extrapolate how that statistic will translate into a measurable outcome that positively affects their organization. This is a significant (and likely erroneous) leap of

faith when it comes to the ability of practitioners to consume scholarly research. It seems clear that the focus of scholarly versus practitioner articles in addition to the approaches they take when discussing validity and utility are contributing to the prevailing scholar-practitioner gap in the area of selection tests.

Implications for Practitioners

An article in *HR Magazine*, “Avoiding Bias in Preemployment Testing” (Minton-Eversole, 2010) includes an interesting discussion on why HR professionals need to become more shrewd and educated consumers of assessments but are unlikely to do so. The author makes the point that “the huge cost and performance ramifications associated with poor employee selection makes this a critical competency for all HR professionals” (Testing Demands section, para. 4). However, quoting Elaine D. Pulakos (COO for the industrial/organizational psychology consulting firm PDRI, a subsidiary of Previsor), the author makes the point that

Many HR professionals have misconceptions about both the value of formal assessments and the types of assessments that have proven to be most effective. This, coupled with the fact that the area of selection testing is inherently technical and difficult to understand has led to an underutilization of formal assessments in organizations. With everything that HR generalists have on their plates, it’s unlikely that they’re going to be able to devote the time necessary to becoming experts in psychological assessments (Testing Demands section, para. 6-7).

Minton-Eversole suggests that practitioners consult with I/O Psychologists who possess the expertise necessary to assist HR in selecting and utilizing hiring assessments.

Others, however, do believe that HR practitioners are currently capable of grasping this technical information about psychological assessments (Wanek, 1999). The *HR Magazine* article “Hire Intelligence” (Roberts, 2011a) discusses the state of the HR profession relative to data-driven decision-making. The author argues that HR has lagged behind in utilizing metrics due to the fact that they were one of the last functions to install enterprise software to collect key data, and, thus, the emphasis on being “strategic” in the HR function (which implies greater rigor in decision-making) is a relatively new phenomenon. “Just a few years ago, HR professionals struggled to understand chi squares and correlation coefficients” (para. 4). Now, however, Roberts states that many HR managers are moving beyond metrics and are actually trained to analyze important data. Therefore, it would be useful for future research to determine the extent to which HR practitioners are actually equipped (or not) to utilize and understand advanced statistical data.

Despite the fact that numerous studies have demonstrated a concrete connection between utilizing more valid screening processes and bottom-line organizational performance (Schmidt & Hunter, 1998; Terpstra & Rozell, 1993; Terpstra & Rozell, 1997), many practitioners still seem to rely on less scientific methods like subjective impressions or intuition (Terpstra & Rozell 1997). Trade publications have a significant opportunity to reach HR professionals and educate them in ways that can make them more successful in their roles, potentially elevating the status of the entire profession and having a positive financial impact on organizations worldwide. In this way, magazine publishers have tremendous power and influence and should consider it a great responsibility to ensure that they are publishing high quality information that is both

practical and research-based. As Rynes and colleagues (2007) point out, “any periodical that aspires to be educational has a social obligation to find ways of differentiating among the strengths of alternative claims” (p. 1002).

Suggestions for practitioners themselves would include increasing their knowledge regarding the predictive validity of different hiring methods via methods such as college courses, business books, or tutoring from consultants. Using this knowledge to create a highly effective hiring strategy for their organization will increase the HR practitioner’s value to the organization and have an impact on the organization’s bottom line. In instances where practitioners are utilizing an assessment instrument, it is recommended that they conduct a concurrent validation study to ensure that test results are correlated with some desired metric or outcome like turnover, productivity, or sales volume to ensure that there is a return-on-investment for the assessment instrument being used.

Implications for Scholars

The acknowledged scholar-practitioner gap should be of concern to scholars as they reflect on what effect researchers are actually having (König, 2010). Rynes and colleagues (2007) made an excellent argument as to why it is critical for scholars to acknowledge and consider practitioner needs when determining their research directions:

For some time now, academic management researchers have been losing ground to consultants (and more recently, journalists [e.g., Friedman, 2006]) as sources of ideas and advice for practitioners and policy makers (Abrahamson & Eisenman, 2001; Bartlett, 2007; Rigby, 2001). Unfortunately, this decline is occurring at the same time that academics’ dependence on practitioners for resources is

increasing (Trank & Rynes, 2003) and global competition and growth are increasing the need for both more effective and more sustainable organizations (Abrahamson & Eisenman, 2001; Bansal & Gao, 2006). Some believe that our failure to “matter more” (Hambrick, 1994) is approaching crisis stage (e.g., Bartlett, 2007) (p. 999).

This study has confirmed that a substantive gap exists in the information that HR professionals receive through practitioner-oriented publications and the information published in academic journals by scholars. Because it is unlikely that all practitioners will choose to (or be required to) receive the statistical education necessary to understand scholarly research findings as they are currently written, the onus is really on scholars to make their findings more accessible to practitioners. Rynes et al. (2002) point out that “unless research is translated for nonacademic audiences, it has little chance of being read and understood by most practicing managers” (p. 164). Perhaps, if nothing else, scholars could improve their skill at writing in “practice-based language” as Deadrick & Gibson (2007, p. 137) recommend as a way to close the gap. Studies also suggest that scholars make attempts to tie the use of predictive selection practices to organizational outcomes rather than just validity data (Anderson et al., 2004; Terpstra & Rozell, 1997). It appears, from the current study, that researchers have not embraced this advice.

Limitations of current study

One possible limitation of the study is that there is a fair amount of researcher judgment and subjectivity inherent to content analysis. In certain analysis activities, like selecting articles for inclusion in the analysis or creating categories for “utility,” it would have been beneficial to have had additional independent coders so that inter-coder

reliability could have been determined. Another limitation is that this study only looked at practitioner literature as a source of information about selection assessments. It is likely that HR practitioners receive information pertinent to selection assessments from other sources as well (i.e., SHRM website, white papers, business books), and it would be worthwhile for future studies to examine to what extent the messages conveyed in those sources are similar to or different than the ones conveyed in the practitioner publications reviewed in this study.

Suggestions for future research

As was previously mentioned, one area for future research would be to determine the degree to which practitioners actually are educated and trained to understand and utilize statistical and psychometric data. This is critical in determining the extent to which scholars and test publishers need to alter their current approaches in presenting these data relative to selection assessments. Confirming the extent to which practitioners are intimidated by the technical complexity of selection instrument psychometrics is important in understanding the best ways to minimize the scholar-practitioner gap.

Future studies should also examine how practitioners make decisions relative to selection assessments. This could potentially be done through survey research that determines the relative weight various factors carry in HR practitioner decisions-making processes when determining whether or not to implement a selection assessment. Some of the potential factors that could be included in the study are sales person persuasiveness, persuasiveness of marketing materials, availability of case studies demonstrating instrument utility, evidence of predictive validity, prevalence of the use of a particular instrument within an industry or community, and legal considerations.

Next, although some studies have attempted to assess the prevalence of publication bias from test publishers, further research could attempt to analyze technical information provided by test publishers to assess the relative strength of this information. For example, were adequate and effective sampling procedures used to determine reliability? What methods were used to determine content and criterion-related validity? Such a study could help determine the extent to which the information provided by test publishers can be relied on by practitioners when deciding whether or not to implement a particular selection instrument. This is important because most (if not all) of the information practitioners need to make selection-assessment-related decisions is provided by sales people and test publishers rather than a neutral party.

Conclusion

“The scientist-practitioner gap in personnel selection is large” (König et al., 2010, p. 99). This study sought to further understand potential causes of this gap by examining the messages that practitioners receive through trade publications relative to selection assessments compared to the research generated by scholars. It was found that publications from each orientation had significantly different approaches, with practitioner articles discussing “big picture” issues that were relevant to the job requirements of HR professionals and scholarly articles tending to address more specific issues that were important to theory-building or construct refinement. This indicates differences in both interests and approach between scholars and practitioners.

Next, there was no overlap between the two orientations in how validity was discussed. Considering the criticality of this topic in relation to selection methods, it is important that ways are found to convey to practitioners the importance of obtaining and

incorporating this data. Furthermore, scholars could take steps to make complex validity information more accessible for practitioners.

Lastly, it is clear that practitioners are interested in understanding how using an assessment instrument is likely to affect them in terms of financial or other organizational outcomes. Scholars, on the other hand, appear to have little to no inclination to explicitly state the benefits of using a predictive measure beyond reporting validity coefficients.

This study uncovered several areas where scholars and practitioners can take steps toward each other to create common dialog in order to bridge the gap.

REFERENCES

- *Aguinis, H., Culpepper, S.A., & Pierce, C. A. (2010). Revival of test bias research in preemployment testing. *Journal of Applied Psychology, 95*, 648-680.
- Anderson, N., Lievens, F., van Dam, K., & Ryan, A. M. (2004). Future perspectives on employee selection: Key directions for future research and practice. *Applied Psychology, 53*, 487-501.
- Aronson, Z. H. & Reilly, R. R. (2006). Personality validity: The role of schemas and motivated reasoning. *International Journal of Selection and Assessment, 14*, 372-380.
- *Arthur, W. Jr., Bell, S. T., Villado, A. J., & Doverspike, D. (2006). The use of person-organization fit in employment decision making: An assessment of its criterion-related validity. *Journal of Applied Psychology, 91*, 786-801.
- Arthur, W. & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology, 93*, 435-442.
- Barrick, M. R. & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta- analysis. *Personnel Psychology, 44*, 1-25.
- Barrick, M. R. & Mount, M. K. (1993). Autonomy as a moderator of the relationship between the big five personality dimensions and job performance. *Journal of Applied Psychology, 78*, 111-118.

- Bernardin, H. J. (1977). The relationship of personality variables to organizational withdrawal. *Personnel Psychology*, 30, 17-27.
- *Berry, C. M., Clark, M. A., & McClure, T. K. (2011). Racial/Ethnic differences in the criterion-related validity of cognitive ability tests: A qualitative and quantitative review. *Journal of Applied Psychology*, 96, 881-906.
- *Berry, C. M., Gruys, M. L., & Sackett, P. R. (2006). Educational attainments as a proxy for cognitive ability in selection: Effects on levels of cognitive ability and adverse impact. *Journal of Applied Psychology*, 91, 696-705.
- *Berry, C. M., Sackett, P. R., & Wiemann, S. (2007). A review of recent developments in integrity test research. *Personnel Psychology*, 60, 271-301.
- *Berry, C. M. & Sackett, P. R. (2009). Faking in personnel selection: Tradeoffs in performance versus fairness resulting from two cut-score strategies. *Personnel Psychology*, 62, 835-863.
- *Bing, M. N., Stewart, S. M., Davison, H. K., Green, P. D., McIntyre, M. D., & James, L. R. (2007). An integrative typology of personality assessment for aggression: Implications for predicting counterproductive workplace behavior. *Journal of Applied Psychology*, 92, 722-744.
- *Bolch, M. (2008a). Nice work. *HR Magazine*, 53(2).
- *Bolch, M. (2008b). Closing the sale. *HR Magazine*, 53(10).
- Braun, V. & Clarke, V. (2006). Using qualitative content analysis in psychology. *Qualitative Research in Psychology*, 3, 77-101.

- Burch, G. J. & Anderson, N. (2004). Measuring person-team fit: Development and validation of the team selection inventory. *Journal of Managerial Psychology, 19*, 406-426.
- Byrne, J. C., Dominick, P. G., Smither, J. W., & Reilly, R. R. (2007). Examination of the discriminant, convergent, and criterion-related validity of self-ratings on the emotional competence inventory. *International Journal of Selection and Assessment, 15*, 341-353.
- Christiansen, N. D., Janovics, J. E., & Siers, B. P. (2010). Emotional intelligence in selection contexts: Measurement method, criterion-related validity, and vulnerability to response distortion. *International Journal of Selection and Assessment, 18*, 87-101.
- Deadrick, D. L. & Gibson, P. A. (2007). An examination of the research-practice gap in HR: Comparing topics of interest to HR academics and HR professionals. *Human Resource Management Review, 17*, 131-139.
- *Dilchert, S., Ones, D. S., Davis, R. D., & Rostow, C. D. (2007). Cognitive ability predicts objectively measured counterproductive work behaviors. *Journal of Applied Psychology, 92*, 616-627.
- *Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology, 91*, 40-57.
- *Ellingson, J. E., Sackett, P. R., & Connelly, B. S. (2007). Personality assessment across selection and development contexts: Insights into response distortion. *Journal of*

Applied Psychology, 92, 386-395.

- *Felton-O'Brien, M. (2008). Screening's worth. *HR Executive*. February 1.
- *Flander, S. (2009a). Recessionary Assessments. *HR Executive*. June 16.
- *Flander, S. (2009b). Testing for team engagement. *HR Executive*. June 16.
- Furnham, A. (2008). HR professionals' beliefs about, and knowledge of, assessment techniques and psychometric tests. *International Journal of Selection and Assessment*, 16, 300-306.
- Furnham, A., Dissou, G., Sloan, P., & Chamorro-Premuzie, T. (2007). Personality and intelligence in business people: A study of two personality and two intelligence measures. *Journal of Business Psychology*, 22, 99-109.
- *Geisinger, K. F. (2008). Screening: Testing the limits. *HR Executive*. November 19.
- *Gill, C. M. & Hodgkinson, G. P. (2007). Development and validation of the five-factor model questionnaire (GGMQ): An adjectival-based personality inventory for use in occupational settings. *Personnel Psychology*, 60, 731-766.
- *Greenstein, L. (2008). Web of deceit. *HR Executive*, June 16.
- *Hausknecht, J. P. (2010). Candidate persistence and personality test practice effects: Implications for staffing system management. *Personnel Psychology*, 63, 299-324.
- *Heggestad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology*, 91, 9-24.
- Heneman, H., & Judge, T. (2008). *Staffing organizations*. New York, NY: McGraw-Hill.

- *Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, 92, 1270-1285.
- Hogan, J. & Hogan, R. (1989). How to measure employee reliability. *Journal of Applied Psychology*, 74, 273-279.
- *Joseph, D. L. & Newman, D. A. (2010). Emotional intelligence: An integrative meta-analysis and cascading model. *Journal of Applied Psychology*, 95, 54-78.
- *Judge, T. A. & Erez, A. (2007). Interaction and intersection: The constellation of emotional stability and extraversion in predicting performance. *Personnel Psychology*, 60, 573-596.
- Judge, T. A., Higgins, C. A., Thorensen, C. J., & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology*, 52, 621-650.
- *Judge, T. A., LePine, J. A., & Rich, B. L. (2006). Loving yourself abundantly: Relationship of the narcissistic personality to self- and other perceptions of workplace deviance, leadership, and task and contextual performance. *Journal of Applied Psychology*, 91, 762-776.
- Judge, T. A., Martocchio, J. J., & Thorensen, C. J. (1997). Five-factor model of personality and employee absence. *Journal of Applied Psychology*, 82, 745-755.
- *Komar, S., Brown, D. J., Komar, J. A., & Robie, C. (2008). Faking and the validity of conscientiousness: A monte carlo investigation. *Journal of Applied Psychology*, 93, 140-154.

- König, C. J., Klehe, U., Berchtold, M., & Kleinmann, M. (2010) Reasons for being selective when choosing personnel selection procedures. *International Journal of Selection and Assessment*, 18, 17-26.
- *Landers, R. N., Sackett, P. R., & Tuzinski, K. A. (2011). Retesting after initial failure, coaching rumors, and warnings against faking in online personality measures for selection. *Journal of Applied Psychology*, 96, 202-210.
- *Le, H., Oh, I., Robbins, S. B., Ilies, R., Holland, E., & Westrick, P. (2011). Too much of a good thing: Curvilinear relationship between personality traits and job performance. *Journal of Applied Psychology*, 96, 113-133.
- *Lievens, F., De Corte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology*, 93, 268-279.
- *Marcus, B., Lee, K., & Ashton, M. C. (2007). Personality dimensions explaining relationships between integrity tests and counterproductive behavior: Big five, or one in addition? *Personnel Psychology*, 60, 1-34.
- *McDaniel, M. A., Rothstein, H. R., & Whetzel, D. L. (2006). Publication bias: A case study of four test vendors. *Personnel Psychology*, 59, 927-953.
- *Minton-Eversole, T. (2010). Avoiding bias in pre-employment testing. *HR Magazine*, 55(12).
- *Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007a). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 60, 683-729.

- *Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007b). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology*, 60, 1029-1049.
- Morreale, S., Backlund, P., Hay, E., & Moore, M. (2011). Assessment of oral communication: A major review of the historical development and trends in the movement from 1975 to 2009. *Communication Education*, 60, 255-278.
- *Mount, M. K., Oh, I., & Burns, M. (2008). Incremental validity of perceptual speed and accuracy over general mental ability. *Personnel Psychology*, 61, 113-139.
- Murphy, K. R. (2000). Impact of assessments of validity generalization and situational specificity on the science and practice of personnel selection. *International Journal of Selection and Assessment*, 8, 194-206.
- *Murphy, K. R., Dzieweczynski, J. L., & Zhang, Y. (2009). Positive manifold limits the relevance of content-matching strategies for validating selection test batteries. *Journal of Applied Psychology*, 94, 1018-1031.
- *O'Brien, M. (2009). Essential Assessments. *HR Executive*, May 16
- *O'Brien, M. (2010). Assessing a threat. *HR Executive*, April 1.
- *Oh, I., Wang, G. & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology*, 96, 762-773.
- *Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology*, 60, 995-1027.

- Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, 78, 679-703.
- Piotrowski, C., & Armstrong, T. (2006). Current recruitment and selection practices: A national survey of fortune 1000 firms. *North American Journal of Psychology*, 8, 489-496.
- *Roberts, B. (2011a). Hire intelligence. *HR Magazine*, 56(5).
- *Roberts, B. (2011b). Your cheating heart. *HR Magazine*, 56(6).
- *Roberts, B. W., Harms, P. D., Caspi, A., & Moffitt, T. E. (2007). Predicting the counterproductive employee in a child-to-adult prospective study. *Journal of Applied Psychology*, 92, 1427-1436.
- Rynes, S. L., Colbert, A. E., & Brown, K. G. (2002). HR professionals' beliefs about effective human resource practices: Correspondence between research and practice. *Human Resource Management*, 41, 149-174.
- Rynes, S. L., Giluk, T. L., & Brown, K. G. (2007). The very separate worlds of academic and practitioner periodicals in human resource management: Implications for evidence-based management. *Academy of Management Journal*, 50, 987-1008.
- Sackett, P. R., Burris, L. R., & Callahan, C. (1989). Integrity testing for personnel selection: An update. *Personnel Psychology*, 42, 491-527.
- Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.

- Schmidt, F. L. & Hunter, J. E. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, 86, 162-173.
- *Schmidt, F. L., Shaffer, J. A., & Oh, I. (2008). Increased accuracy for range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology*, 61, 827-868.
- *Schmitt, N. & Oswald, F. L. (2006). The impact of corrections for faking on the validity of noncognitive measures in selection settings. *Journal of Applied Psychology*, 91, 613-621.
- Scroggins, W. A., Thomas, S. L., & Morris, J. A. (2008). Psychological testing in personnel selection, part II: The refinement of methods and standards in employee selection. *Public Personnel Management*, 37, 185-195.
- Scroggins, W. A., Thomas, S. L., & Morris, J. A. (2009). Psychological testing in personnel selection, part III: The resurgence of personality testing. *Public Personnel Management*, 38, 67-74.
- *Stark, S., Chernyshenko, O. S., Drasgow, F., & Williams, B. A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology*, 91, 25-39.
- Terpstra, D. E. & Rozell, E. J. (1993). The relationship of staffing practices to organizational level measures of performance. *Personnel Psychology*, 46, 27-48.
- Terpstra, D. E. & Rozell, E. J. (1997). Why some potentially effective staffing practices are seldom used. *Public Personnel Management*, 26, 483-495.

- *Tett, R. P. & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology*, 60, 967-993.
- Tsaousis, I., & Nikolaou, I. (2001). The stability of the five-factor model of personality in personnel selection and assessment in Greece. *International Journal of Selection and Assessment*, 9, 290-301.
- U.S. Department of Labor (1999). Testing and assessment: An employer's guide to good practices. Washington DC: Employment and Training Administration.
- *Van Iddekinge, C. H., Putka, D. J., & Campbell, J. P. (2011). Reconsidering vocational interest for personnel selection: The validity of an interest-based selection test in relationship to job knowledge, job performance, and continuance intentions. *Journal of Applied Psychology*, 96, 13-33.
- Van Rooy, D. L., Viswesvaran, C., & Pluta, P. (2005). An evaluation of construct validity: What is this thing called emotional intelligence? *Human Performance*, 18, 445-462.
- Viswesvaran, C., Deller, J., & Ones, D. S. (2007). Personality measures in personnel selection: Some new contributions. *International Journal of Selection and Assessment*, 15, 354-358.
- Wanek, J. E. (1999). Integrity and honesty testing: What do we know? How do we use it? *International Journal of Selection and Assessment*, 7, 183-195.
- Werner, S. H., Jones, J. W., & Steffy, B. D. (1989). The relationship between intelligence, honesty, and theft admissions. *Educational and Psychological Measurement*, 49, 921-927.

- *Wu, J. & Lebreton, J. M. (2011). Reconsidering the dispositional basis of counterproductive work behavior: The role of aberrant personality. *Personnel Psychology, 64*, 593-626.
- Zhang, Y., & Wildemuth, B. M. (2009). Qualitative analysis of content. In B. Wildemuth (Ed.), *Applications of Social Research Methods to Questions in Information and Library Science* (pp.308-319). Westport, CT
- Zickar, M. J. & Highhouse, S. (2001). Measuring prestige of journals in industrial-organizational psychology. *The Industrial-Organizational Psychologist, 38*.
- *Zielinski, D. (2011). Effective Assessments. *HR Magazine, 56*(1).
- * *Articles marked with an asterisk were included in the qualitative content analysis.*

APPENDIX

This appendix includes the raw data extracted from each article relative to each of the five research questions that was subsequently content analyzed. The five research questions were as follows:

RQ 1: What was the main purpose of the article?

RQ 2: What types of assessments were discussed?

RQ 3: Were any commercially available assessment instruments mentioned specifically? If so, which one(s)?

RQ 4: How was validity discussed?

RQ 5: How was utility discussed?

The 49 articles included in content analysis are listed below, alphabetical by author, organized by source.

Practitioner, *HR Executive*

Felton-O'Brien, M. (2008). Screening's worth. *HR Executive*. February 1.

RQ 1: Return-on-investment of assessments (case study meta-analysis done by Previsor)

RQ 2: "conscientiousness"

RQ 3: Personnel Decisions International (PDI), Development Decisions International (DDI), Previsor

RQ 4: N/A

RQ 5: highest scorers generated \$1.3 million more per month in sales revenue.

High scores on conscientiousness resulted in \$50K less in stock shrinkage per year which translated to \$78 million in company-wide savings.

Flander, S. (2009a). Recessionary Assessments. *HR Executive*. June 16.

RQ 1: Increased need for/use of assessments due to recession

RQ 2: Problem solving skills, critical thinking, personality, GMA, culture fit/values

RQ 3: PDI, Kenexa, Pearson's Talent Lens group, DDI, Previsor, APT Inc.

RQ 4: Help you increase the probability of making the right selection; bring some science into what can be a subjective process of looking for cultural fit.

RQ 5: "predictive validation study" found low scorers were four times more likely to leave within 90 days than high scorers, and high scorers were 15% more productive; Recession means that companies have to be more competitive than ever, applicant pool is bigger and of higher quality, which means new hires have to be stars who can hit the ground running (because the company doesn't have the luxury of waiting), who stay with the company (because companies can't afford to replace them)... like "finding diamonds in a haystack")

Flander, S. (2009b). Testing for team engagement. *HR Executive*. June 16.

RQ 1: Screen for team engagement-raising characteristics

RQ 2: Personality

RQ 3: Gallup

RQ 4: N/A

RQ 5: N/A

Geisinger, K.F. (2008). Screening: Testing the limits. *HR Executive*. November 19.

RQ 1: How technology advances, customer demands, and internationalization have impacted screening processes

RQ 2: GMA, personality, computer skills tests, certifications, electronic in-basket, work samples

RQ 3: N/A

RQ 4: When researchers build tests, they conduct research to demonstrate the validity of the tests (that is, the accuracy with which the test predicts future job performance); As tests become increasingly proprietary, test publishers are more secretive regarding the nature and usefulness of measures. Information isn't always provided to test users. If we don't know to what extent measures actually are effective, we lose our scientific basis; Personality test lack the widespread validation success that one finds in ability and work sample measures. They may be useful in some situations but demand hard data/empirical research indicating job relevance.

RQ 5: Utility is limited for personality tests—expect use to decline because they simply do not predict performance; The very nature of the measures we use is changing due to evidence supporting their use; Useful when you have a big applicant pool and need the best fitting candidate—should rely heavily on proper use of personnel testing.

Greenstein, L. (2008). Web of deceit. *HR Executive*, June 16.

RQ 1: Impact of faking/cheating on assessments

RQ 2: Computer knowledge/skills tests, behavioral, cognitive

RQ 3: Select International, SHL, Previsor, Kenexa

RQ 4: Validate assessment results through behavioral interviews, reference checks, and background screens.

RQ 5: N/A

O'Brien, M. (2009). Essential Assessments. *HR Executive*, May 16

RQ 1: Use assessments to identify “high risk” individuals (reliability, safety, etc.)

RQ 2: Integrity

RQ 3: Profiles International’s SOS II, Insight Worldwide’s Quickstaff, SHL’s Dependability and Safety Instrument

RQ 4: N/A

RQ 5: Cost of unreliable hires according to U.S. Bureau of Labor statistics: companies lose about \$2.8 million work days per year from absenteeism; manage risk; objective data to protect from discriminatory claims.

O'Brien, M. (2010). Assessing a threat. *HR Executive*, April 1.

RQ 1: How to detect potential violent offenders

RQ 2: Integrity

RQ 3: Israel-based suspect detection system COGITO 1003

RQ 4: Detects hostile intents of assailants with remarkable degree of accuracy

RQ 5: Countless workers could be spared dealing with ticking time bombs (i.e.: Fort Hood, and Alabama Huntsville professor shootings)

Practitioner, HR Magazine

Bolch, M. (2008a). Nice work. *HR Magazine*, 53(2).

RQ 1: Screening for “nice”

RQ 2: Personality

RQ 3: DDI, Kenexa, Previsor, Infinet Assessments Inc.

RQ 4: Previsor is a provider of validated employee screening products; Testing is one of the most predictive parts of the application process (resume results are approximately zero predictive, interview predictiveness varies widely); Infinet can create, validate, and roll out custom assessments in four weeks; Each competency is weighted according to its predictive value; Off the shelf assessments validated for certain jobs may cost less but be less valuable than custom tools; Cost of validation study that will withstand legal scrutiny costs about 5-20K.

RQ 5: Cut recruiting time and improve hiring success; turnover dropped by two-thirds; highest scorers had 24% increase in sales rates; have looked at return-on-investment and believe it will happen through reduction of first year turnover.

Bolch, M. (2008b). Closing the sale. *HR Magazine*, 53(10).

RQ 1: Review of best methods for hiring sales people

RQ 2: personality, cognitive ability

RQ 3: Birkman, Caliper, Hogan, PreVisor, Strong Interest Inventory

RQ 4: Instinctive feelings backed up by validated assessments; Previsor conducted a validations study.

RQ 5: Reduced turnover, productivity hike; better scorers signed eleven times more clients per month, selling \$434 more in features and accessories to compliment original sale; Benefits include retention and higher profits.

Minton-Eversole, T. (2010). Avoiding bias in pre-employment testing. *HR Magazine*, 55(12).

RQ1: How to assess test bias

RQ2: GMA

RQ3: Previsor, Selection Research International (SRI)

RQ4: As long as assessments have been validated and proven to be without bias, such testing is acceptable; Show job relatedness through professionally sound validation study; Strongest validation evidence is obtained by demonstrating that people who score higher on tests actually perform better on the job.

RQ5: Good assessment instruments are the most effective and efficient tools HR have available to make hiring decisions; Have been repeatedly shown to yield important bottom-line outcomes such as increased productivity, reduced turnover, enhanced employee engagement and manager satisfaction, among others; Add tremendous value.

Roberts, B. (2011a). Hire intelligence. *HR Magazine*, 56(5).

RQ 1: Data driven decision making/metrics

RQ 2: Skills, knowledge, attitude, behavior, cognitive ability

RQ 3: Kenexa, People Answers, Success Factors, Knowledge Infusion

RQ 4: Data-driven decision-making; In developing a custom assessment, a sample of at least 300 is needed for statistical validity; validated and checked for adverse impact; sample population answered about 250 questions from Kenexa's library, analyzed to see which traits correlated most highly with performance and tenure;

Revalidated and fine tuned the test; Cognitive ability is most closely correlated with turnover and sales.

RQ 5: Lower turnover, increased sales per hour, cut turnover in half.

Roberts, B. (2011b). Your cheating heart. *HR Magazine*, 56(6).

RQ 1: Overview of integrity tests

RQ 2: Integrity, personality, situational judgment

RQ 3: Merchants Information Solutions, Previsor, Skill Survey Inc., Success Performance Solutions, Vangent, Wonderlic

RQ 4: Test publishers, academics, and employers have found integrity tests to be valid measures of counterproductive behaviors; Provide measurable results; Have been proved in statistical analyses to be predictive of theft and other negative behaviors on the job; More broad based personality tests which are also valid and shed light on many character traits including integrity; By any measure—scientific, legal, return-on-investment—integrity testing gets at least a passing grade; integrity tests have been validated repeatedly by developers, independent researchers, and HR professionals; Ones' meta-analysis found that integrity tests are statistically valid and predict what they purport to; Among all types of pre-hire assessments, integrity tests have the highest validity in predicting undesirable behaviors at work; Not all tests are equal—make sure vendors have evidence of validation; All assessments—interviews, resume screenings, other tests—have false positives and negatives. Higher validity means lower error rate. Using integrity tests minimizes false positives/negatives compared to other methods of employee selection.

RQ 5: Reduced workers compensations claims; The most important results are the business benefits, i.e.: better sales and customer feedback; Personality tests resulted in increased performance reviews, lower turnover, and increased sales; Average workers compensation claim size was \$3,466 (unscreened) compared to \$2,119 (screened). This is statistically significant (greater than expected due to chance); Return-on-investment of 50% in one year (means half the cost of the tests were recovered in savings from lower workers compensation claims); Another study found a return-on-investment of 846%; Integrity tests provide predictions regarding where individuals are likely to fall along a negative behavior continuum. This knowledge allows HR to manage risk at the level they're comfortable with.

Zielinski, D. (2011). Effective Assessments. *HR Magazine*, 56(1).

RQ 1: Evolution of assessments based on technology

RQ 2: Knowledge, cultural fit, "risk," skills

RQ 3: DDI, Select International, Shakers Virtual Job Tryout Simulation, PDI

RQ 4: Questions about validity to ask vendors—when were validity tests done?

For what jobs was the instrument validated? What were the results of the validity test? What was the test coefficient? What was the sample size? Was the test shown to be reliable as well as valid?; Properly validated assessments have long been viewed as an important arrow in the performance management quiver.

RQ 5: \$1.7 million in cost savings in teller turnover in one year; whether the goal is improving cost of hire or quality of hire, assessments have become an indispensable piece of the decision-making puzzle.

Scholarly, Journal of Applied Psychology

Aguinis, H., Culpepper, S.A., & Pierce, C.A. (2010). Revival of test bias research in preemployment testing. *Journal of Applied Psychology*, 95(4), 648-680.

RQ1: Reevaluating test bias in preemployment testing

RQ2: GMA

RQ3: N/A

RQ4: Validity coefficients reported

RQ5: N/A

Arthur, W. Jr., Bell, S.T., Villado, A.J., & Doverspike, D. (2006). The use of person-organization fit in employment decision making: An assessment of its criterion-related validity. *Journal of Applied Psychology*, 91(4), 786-801.

RQ 1: Investigate criterion related validity of P-O fit as a predictor of job performance and turnover

RQ 2: Personality (P-O fit)

RQ 3: N/A

RQ 4: Validity coefficients reported

RQ 5: Assumed

Berry, C.M., Clark, M.A., & McClure, T.K. (2011). Racial/Ethnic differences in the criterion-related validity of cognitive ability tests: A qualitative and quantitative review. *Journal of Applied Psychology*, 96(5), 881-906.

RQ 1: To look at differences between ethnicities on GMA (bias)

RQ 2: GMA

RQ 3: General Aptitude Test Battery, Wonderlic

RQ 4: Validity coefficients reported

RQ 5: Because there are high stakes surrounding the use of cognitive ability tests for selection and placement, the validity of the inferences made from these tests is of paramount importance; the differences in validity of tests for different groups affects utility. Utility is a function of the validity of a test. Other factors influencing utility are average predictor scores, standard deviation of performance; Utility is measured by output, dollars, mean performance.

Berry, C.M., Gruys, M.L., & Sackett, P.R. (2006). Educational attainments as a proxy for cognitive ability in selection: Effects on levels of cognitive ability and adverse impact. *Journal of Applied Psychology*, 91(3), 696-705.

RQ 1: To see how screening for educational attainment compared to screening directly for GMA

RQ 2: Cognitive Ability

RQ 3: Armed Forces Qualification Test (AFQT)

RQ 4: Validity coefficients reported

RQ 5: Cognitive ability has long been demonstrated to be an important determinant of performance and training across a broad spectrum of jobs; Subsequent loss of criterion related validity if choosing educational attainment over GMA may not be worth it to some companies.

Bing, M.N., Stewart, S.M., Davison, H.K., Green, P.D., McIntyre, M.D., & James, L.R. (2007). An integrative typology of personality assessment for aggression: Implications for predicting counterproductive workplace behavior. *Journal of Applied Psychology*, 92(3), 722-744.

RQ 1: To test a process for measuring aggression

RQ 2: Personality

RQ 3: N/A

RQ 4: Validity coefficients reported

RQ 5: Workplace complaints and grievances can have serious financial and wellness consequences for both employees and the company.

Dilchert, S., Ones, D.S., Davis, R.D., & Rostow, C.D. (2007). Cognitive ability predicts objectively measured counterproductive work behaviors. *Journal of Applied Psychology*, 92(3), 616-627.

RQ 1: To examine the relationship between GMA and counterproductive workplace behaviors (CWB)

RQ 2: GMA

RQ 3: Shipley Institute of Living Scale (SILS)

RQ 4: Validity coefficients reported

RQ 5: CWB results in exorbitant costs to organizations, typically ranging in the billions of dollars; Since GMA predicts CWB, need to factor into the utility equation documented savings from reduced CWB.

Dudley, N.M., Orvis, K.A., Lebiecki, J.E., & Cortina, J.M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology*, 91(1), 40-57.

RQ 1: To see if narrow traits predict job performance above and beyond global conscientiousness

RQ 2: Personality

RQ 3: N/A

RQ 4: Validity coefficients reported

RQ 5: Narrow traits are beneficial for prediction, but their value for personnel selection requires a careful match of the particular narrow trait to the occupation and job performance criterion in question.

Ellingson, J.E., Sackett, P.R., & Connelly, B.S. (2007). Personality assessment across selection and development contexts: Insights into response distortion. *Journal of Applied Psychology*, 92(2), 386-395.

RQ 1: Faking in selection versus development contexts

RQ 2: Personality

RQ 3: California Psychological Inventory (CPI)

RQ 4: Validity coefficients reported

RQ 5: N/A

Heggestad, E.D., Morrison, M., Reeve, C.L., & McCloy, R.A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology*, 91(1), 9-24.

RQ 1: To see if alternate scale is more resistant to faking

RQ 2: Personality

RQ 3: NEO Big Five

RQ 4: Validity coefficients reported

RQ 5: Evidence of criterion related validity has led to optimism for the use of personality testing on personnel selection contexts.

Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, 92(5), 1270-1285.

RQ 1: Measuring faking

RQ 2: Personality

RQ 3: Hogan Personality Inventory (HPI)

RQ 4: Validity coefficients reported

RQ 5: N/A

Judge, T.A., LePine, J.A., & Rich, B.L. (2006). Loving yourself abundantly:

Relationship of the narcissistic personality to self- and other perceptions of workplace deviance, leadership, and task and contextual performance. *Journal of Applied Psychology*, 91(4), 762-776.

RQ 1: To see if narcissism predicts job performance

RQ 2: Personality

RQ 3: Narcissistic Personality Inventory, NEO-FFI, Leadership Practices

RQ 4: Validity coefficients reported

RQ 5: N/A

Joseph, D.L. & Newman, D.A. (2010). Emotional intelligence: An integrative meta-analysis and cascading model. *Journal of Applied Psychology*, 95(1), 54-78.

RQ 1: To propose a new model of EI that incorporates GMA and personality and test it and other hypotheses related to EI

RQ 2: Personality, GMA

RQ 3: MSCEIT, MEIS, WLEIS, EIS, and WEIP (EI “ability model” tools)

RQ 4: Validity coefficients reported

RQ 5: Purported relationship between EI and job performance; Evaluated the potential for EI measures to incrementally predict job performance.

Komar, S., Brown, D.J., Komar, J.A., & Robie, C. (2008). Faking and the validity of conscientiousness: A monte carlo investigation. *Journal of Applied Psychology*, 93(1), 140-154.

RQ 1: Effects of faking

RQ 2: Personality

RQ 3: NEO

RQ 4: Validity coefficients reported

RQ 5: Conscientiousness may be the single best predictor of work performance.

Landers, R.N., Sackett, P.R., & Tuzinski, K.A. (2011). Retesting after initial failure, coaching rumors, and warnings against faking in online personality measures for selection. *Journal of Applied Psychology*, 96(1), 202-210.

RQ 1: To look at ways to address blatant faking

RQ 2: Personality

RQ 3: N/A

RQ 4: Validity coefficients reported

RQ 5: N/A

Le, H., Oh, I., Robbins, S.B., Ilies, R., Holland, E., & Westrick, P. (2011). Too much of a good thing: Curvilinear relationship between personality traits and job performance. *Journal of Applied Psychology*, 96(1), 113-133.

RQ 1: To assess the curvilinear relationship between personality and job performance

RQ 2: Personality

RQ 3: Talent Assessment (developed by ACT for selection based on Big 5)

RQ 4: Validity coefficients reported

RQ 5: No additional cost associated with increased validity—from a utility perspective, an increase in efficiency can reduce expenses and increase savings or profits over time.

Lievens, F., De Corte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology*, 93(2), 268-279.

RQ 1: To see if providing the test taker a “context” increases the validity of personality tests

RQ 2: Personality

RQ 3: NEO

RQ 4: Validity coefficients reported

RQ 5: N/A

Murphy, K.R., Dzieweczynski, J.L., & Zhang, Y. (2009). Positive manifold limits the relevance of content-matching strategies for validating selection test batteries. *Journal of Applied Psychology*, 94(4), 1018-1031.

RQ 1: Don’t necessarily need to match content of tests to content of jobs for certain batteries to be effective predictors

RQ 2: GMA, Personality, Situational Judgment, Biodata

RQ 3: N/A

RQ 4: Validity coefficients reported

RQ 5: Consider the costs of testing (both financial and social costs)—the benefits of these alternatives might not vary as much as their potential costs.

Oh, I., Wang, G. & Mount, M.K. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology*, 96(4), 762-773.

RQ 1: To see if observer ratings are more valid than paper-and-pencil/self-report measures of personality

RQ 2: Personality

RQ 3: N/A

RQ 4: Validity coefficients reported

RQ 5: N/A

Roberts, B.W., Harms, P.D., Caspi, A., & Moffitt, T.E. (2007). Predicting the counterproductive employee in a child-to-adult prospective study. *Journal of Applied Psychology*, 92(5), 1427-1436.

RQ 1: To see if childhood conduct disorder, GMA, and personality predict CWB

RQ 2: Personality, GMA

RQ 3: Multidimensional Personality Questionnaire

RQ 4: Validity coefficients reported

RQ 5: Clearly employers would prefer not to hire employees with a propensity to commit crimes.

Schmitt, N. & Oswald, F.L. (2006). The impact of corrections for faking on the validity of noncognitive measures in selection settings. *Journal of Applied Psychology*, 91(3), 613-621.

RQ 1: Impact of faking

RQ 2: Personality

RQ 3: MMPI, 16PF

RQ 4: Validity coefficients reported

RQ 5: Regarding utility of selection procedures, the most important determinants of mean performance are test validity and selection ratio

Stark, S., Chernyshenko, O.S., Drasgow, F., & Williams, B.A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology*, 91(1), 25-39.

RQ 1: To see if alternates to Likert scales on personality tests could be preferable

RQ 2: Personality

RQ 3: 16PF, NEO, Jackson, Hogan, California Personality Inventory, MMPI

RQ 4: Validity coefficients reported

RQ 5: Changing scoring could change rank ordering of high-scoring individuals which could undermine the utility of personality measures in applied settings; would not dramatically affect predictive validity coefficients but would tend to reduce the utility of personality measures in job selection.

Van Iddekinge, C.H., Putka, D.J., & Campbell, J.P. (2011). Reconsidering vocational interest for personnel selection: The validity of an interest-based selection test in relationship to job knowledge, job performance, and continuance intentions. *Journal of Applied Psychology*, 96(1), 13-33.

RQ 1: Interests as a predictor of job knowledge, performance, and continuance/retention

RQ 2: Interest

RQ 3: Armed Forces Qualification Test (GMA), Work Suitability Inventory (personality), Strong Interest Inventory, Kuder Preference record, Self-Directed Search Interest Inventory, Courses Interest Inventory, Ramak Interest Inventory, AVOICE

RQ 4: Validity coefficients reported

RQ 5: People whose interests are congruent with a given work environment will be more likely to enter that environment, be more satisfied on that environment, perform better and choose to stay longer than people whose interests are incongruent with that environment.

Scholarly, Personnel Psychology

Berry, C.M., Sackett, P.R., & Wiemann, S. (2007). A review of recent developments in integrity test research. *Personnel Psychology*, 60, 271-301.

RQ 1: Summarize new findings regarding integrity testing

RQ 2: Integrity, personality

RQ 3: Personnel Selection Inventory (PSI), Reid Report, Stanton Survey,
California Psychological Inventory (CPI), Personnel Reaction Blank, PDI
Employment Inventory (PDI-EI), Reliability Scale of Hogan Personality Series

RQ 4: Validity coefficients reported

RQ 5: N/A

Berry, C.M. & Sackett, P.R. (2009). Faking in personnel selection: Tradeoffs in
performance versus fairness resulting from two cut-score strategies. *Personnel
Psychology*, 62, 835-863.

RQ 1: To look at the trade-off between predictability and fairness

RQ 2: Personality

RQ 3: N/A

RQ 4: Validity coefficients reported

RQ 5: Faking can have an impact on validity, which impact utility

Gill, C.M. & Hodgkinson, G.P. (2007). Development and validation of the five-factor
model questionnaire (GGMQ): An adjectival-based personality inventory for use
in occupational settings. *Personnel Psychology*, 60, 731-766.

RQ 1: Developing and new personality inventory

RQ 2: Personality

RQ 3: Personality Research Form, Adjective Checklist, Edwards Personal
Preference Schedule, MBTI, CPI, Comrey Personality Scale, MMPI-PD, 16PF,
NEO-PI-R, Hogan HPI

RQ 4: Validity coefficients reported

RQ 5: N/A

Hausknecht, J.P. (2010). Candidate persistence and personality test practice effects:

Implications for staffing system management. *Personnel Psychology*, 63, 299-324.

RQ 1: Effects of retesting

RQ 2: Personality

RQ 3: Hogan PI, California PI, GPPI

RQ 4: Validity coefficients reported

RQ 5: N/A

Judge, T.A. & Erez, A. (2007). Interaction and intersection: The constellation of emotional stability and extraversion in predicting performance. *Personnel Psychology*, 60, 573-596.

RQ 1: To see if the combination of emotional stability and extraversion predicts job performance

RQ 2: Personality

RQ 3: Big Five Inventory

RQ 4: Validity coefficients reported

RQ 5: N/A

Marcus, B., Lee, K., & Ashton, M.C. (2007). Personality dimensions explaining relationships between integrity tests and counterproductive behavior: Big five, or one in addition? *Personnel Psychology*, 60, 1-34.

RQ 1: What personality constructs specifically predict CWB (what underlying constructs are integrity tests actually measuring?)

RQ 2: Integrity

RQ 3: IBES (German integrity test), CPI, Workplace Behavior Questionnaire,
Inventory of Counterproductive Behavior

RQ 4: Validity coefficients reported

RQ 5: N/A

McDaniel, M.A., Rothstein, H.R., & Whetzel, D.L. (2006). Publication bias: A case
study of four test vendors. *Personnel Psychology*, 59, 927-953.

RQ 1: Analyze publication bias

RQ 2: Vaguely addresses personality, GMA, and Integrity

RQ 3: N/A

RQ 4: Validity coefficients reported

RQ 5: N/A

Morgeson, F.P., Campion, M.A., Dipboye, R.L., Hollenbeck, J.R., Murphy, K., &
Schmitt, N. (2007a). Reconsidering the use of personality tests in personnel
selection contexts. *Personnel Psychology*, 60, 683-729.

RQ 1: Review of panel discussion on personality assessments

RQ 2: Personality

RQ 3: N/A

RQ 4: Validity coefficients reported

RQ 5: Implies that personality tests are limited in their ability to predict
performance

Morgeson, F.P., Campion, M.A., Dipboye, R.L., Hollenbeck, J.R., Murphy, K., &

Schmitt, N. (2007b). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology*, 60, 1029-1049.

RQ 1: Effectiveness of personality testing for selection

RQ 2: Personality

RQ 3: 16PF, NEO PI, CPI, Personal Preferences Inventory

RQ 4: Validity coefficients reported

RQ 5: In addition to validity, utility includes the usefulness or value of a selection procedures including the legal defensibility, cost of purchasing and administering the procedure, process flexibility, alignment wit diversity and affirmative action goals, candidate flow statistics, selection ratios, and cycle time to fill a position.

Mount, M.K., Oh, I., & Burns, M. (2008). Incremental validity of perceptual speed and accuracy over general mental ability. *Personnel Psychology*, 61, 113-139.

RQ 1: To see if specific abilities offer incremental validity over GMA

RQ 2: Cognitive Ability and Personality

RQ 3: Wonderlic, Personal Characteristics Inventory (PCI)

RQ 4: Validity coefficients reported

RQ 5: The predictive validity of a selection method is an important attribute because it is a direct determinant of the economic value added.

Ones, D.S., Dilchert, S., Viswesvaran, C., & Judge, T.A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology*, 60, 995-1027.

RQ 1: Review of major findings regarding personality for predicting job performance (reaction to Morgensen et al)

RQ 2: Personality

RQ 3: CPI

RQ 4: Validity coefficients reported

RQ 5: N/A

Schmidt, F.L., Shaffer, J.A., & Oh, I. (2008). Increased accuracy for range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology*, 61, 827-868.

RQ 1: Re-look at validity of personality and GMA using new statistical process for correcting range restriction

RQ 2: Personality and GMA

RQ 3: N/A

RQ 4: Validity coefficients reported

RQ 5: Important practical implications in terms of the utility (practical value) of employment selection systems that incorporate personality and mental ability measures.

Tett, R.P. & Christiansen, N.D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology*, 60, 967-993.

RQ 1: Reaction to Morgeson et al—defending personality/review of research

RQ 2: Personality

RQ 3: NEO-PIR, Hogan Personality Inventory, Jackson Personality Inventory

RQ 4: Validity coefficients reported

RQ 5: Organizations looking to capitalize on the utility of personality tests (their validity comes at relatively low costs).

Wu, J. & Lebreton, J.M. (2011). Reconsidering the dispositional basis of counterproductive work behavior: The role of aberrant personality. *Personnel Psychology*, 64, 593-626.

RQ 1: To see how narcissism, Machiavellianism, and Psychopathy predict CWB

RQ 2: Personality

RQ 3: NEO PI-R, Hogan Development Survey (HDS)

RQ 4: Validity coefficients reported

RQ 5: White collar scandals lead people to ask what could have predicted these.

CURRICULUM VITAE

NAME: Whitney L. Martin

ADDRESS: 144 Fenley Ave
Louisville, KY 40206

DOB: Louisville, KY – November 1, 1977

EDUCATION
& TRAINING: B.A., Communications
University of Kentucky
1996-1999

HONORS: Honors Program, Dean's List, and Merit Scholarship Recipient
1996-1998
Vice President, Communications Student Association
1998
Dickerson Scholarship Recipient from School of Journalism
1997

PRESENTATIONS: Southern States Communication Association Undergraduate
Honors Conference
1999
Kentucky SHRM Conference
2011