


10-31-2016

The Cross-National Generalizability of Biographical Data: An Examination within a Multinational Organization

Adam J. Ducey

University of South Florida, adamducey@gmail.com

Follow this and additional works at: <http://scholarcommons.usf.edu/etd>

 Part of the [Business Administration, Management, and Operations Commons](#), and the [Psychology Commons](#)

Scholar Commons Citation

Ducey, Adam J., "The Cross-National Generalizability of Biographical Data: An Examination within a Multinational Organization" (2016). *Graduate Theses and Dissertations*.
<http://scholarcommons.usf.edu/etd/6493>

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

The Cross-National Generalizability of Biographical Data: An Examination within a
Multinational Organization

by

Adam J. Ducey

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Psychology
College of Arts and Sciences
University of South Florida

Major Professor: Michael D. Covert, Ph.D.
Michael T. Brannick, Ph.D.
Judith B. Bryant, Ph.D.
Chad Dubé, Ph.D.
Stephen Stark, Ph.D.
William H. Young, Ph.D.

Date of Approval:
October 28, 2016

Keywords: personnel selection, biographical data, biodata, cross-national, test validation

Copyright © 2016, Adam J. Ducey

Table of Contents

List of Tables	iii
List of Figures	iv
Abstract	v
Introduction.....	1
The Strategic Importance of Personnel Selection	4
Personnel Selection in Multinational Organizations.....	9
Multinational personnel selection strategies and work analysis.	9
The challenges of global selection systems.	15
The importance of culture for multinational personnel selection.	20
The benefits of global selection systems.	24
The Cross-National Generalizability of Personnel Selection Instruments	26
Cognitive ability and personality.	28
Person-organization fit.....	31
Situational judgment tests.	32
Biographical data.	33
A Review of Biographical Data	35
What is biodata?.....	35
Benefits of biographical data.	38
Theoretical foundation of biodata.	40
Development approaches.	41
Scoring approaches.	43
Conclusion	48
Method - Study 1	50
Participants.....	50
Measures	51
Job performance.....	51
Biographical data inventory.....	52
Procedure	54
Results - Study 1	60
Descriptive Statistics.....	60
Hypothesis Testing.....	66
Method - Study 2	82
Participants.....	82

Measures	83
Job performance.....	83
Biographical data inventory.....	83
Procedure	85
Results - Study 2	89
Descriptive Statistics.....	89
Hypothesis Testing.....	90
Discussion	108
Limitations	112
Future Directions	115
Conclusion	116
References.....	118

List of Tables

Table 1 - Study 1 Demographic Information by Country Cluster	56
Table 2 - Study 1 Descriptive Statistics by Country Cluster	62
Table 3 - Study 1 Biographical Data Inventory Criterion-Related Validity by Country Cluster	75
Table 4 - Study 1 Homogeneity of Error Variance between U.S. and Each Country Cluster	76
Table 5 - Study 1 Differential Prediction between U.S. and Each Country Cluster.....	77
Table 6 - Study 2 Demographic Information by Country Cluster	87
Table 7 - Study 2 Descriptive Statistics by Country Cluster	91
Table 8 - Study 2 Biographical Data Inventory Criterion-Related Validity by Country Cluster	101
Table 9 - Study 2 Homogeneity of Error Variance between U.S. and Each Country Cluster	102
Table 10 - Study 2 Differential Prediction between U.S. and Each Country Cluster.....	103

List of Figures

Figure 1 - Study 1 Distribution of Variables by Country Cluster.....	63
Figure 2 - Study 1 Scatterplot of Criterion-Related Validity by Country Cluster	78
Figure 3 - Study 2 Distribution of Variables by Country Cluster.....	92
Figure 4 - Study 2 Scatterplot of Criterion-Related Validity by Country Cluster	104

Abstract

In an increasingly interconnected economy, organizations are frequently operating beyond national borders. International partnerships, joint ventures, mergers, and acquisitions have expanded the labor market from a domestic to an international perspective. In this environment, multinational organizations demand cost-effective personnel selection tools to help them identify top talent from different countries, geographical regions, and cultures.

The purpose of the current research was to evaluate the global utility of biographical data inventories, a standardized self-report selection method that asks job applicants questions about prior behaviors and experiences. Results from two studies involving participants from 7 country clusters, across four continents, and two different occupational groups, managers and manufacturing technicians, provided evidence to support the hypothesis that the validity of biographical data inventories, empirically keyed in the United States, generalizes to all country clusters examined.

These results are important because they suggest that multinational organizations interested in deploying a single standardized selection system across geographical boundaries may want to consider including biographical data inventories, in addition to other commonly used instruments such as cognitive ability and personality assessments, to enhance the overall validity of their selection systems. This approach has the potential to reduce organizations' costs related to developing, implementing, and maintaining selection processes while enabling them to manage their human capital efficiently by ensuring all new hires have the necessary knowledge, skills, and abilities to succeed on the job and contribute to their strategic objectives.

Introduction

Personnel selection is a core topic in the study of workplace behavior. It is concerned with identifying applicants who possess the necessary knowledge, skills, abilities, and other characteristics (KSAOs) to be effective in a specific job. Organizations commonly use a variety of tools, such as interviews, general mental ability (GMA) tests, and personality inventories, to assess applicants on job-relevant individual differences and make inferences about who is likely to be the best fit for the position in question (Ryan, McFarland, Baron, & Page, 1999; Salgado, Viswesvaran, & Ones, 2001). In an increasingly global economy, the labor market has expanded from a domestic to an international perspective. In this environment, multinational organizations must consider how to implement efficient personnel selection systems on a worldwide scale so they can hire top talent from different geographical regions. When organizations deploy a selection system in an international context, an important question is the extent to which a selection tool is equally effective when implemented in other countries, regions, or cultures (Herriot & Anderson 1997; International Test Commission, 2005; Lievens, 2008; Ployhart & Weekly, 2015; Ryan & Tippins, 2009; Ryan, Wiechmann, & Hemingway, 2003).

The question of a selection tool's effectiveness in a new context is focused on whether or not the instrument's validity is situationally specific or generalizable to a new setting (Schmidt & Hunter, 1977). There is strong meta-analytic evidence to suggest that the validity of commonly used selection instruments generalizes across different jobs and organizations (Schmidt & Hunter, 1998). Little research, however, has examined this question in countries outside of North America (Lievens, 2008; Shen, Kiger, Davies, Rasch, Simon, & Ones, 2011). Since national

social, cultural, and legislative differences influence employees' values, beliefs, and behaviors (Erez, 2011; House, Hanges, Javidan, Dorfman, & Gupta, 2004), it is important to examine if selection instruments have generalized validity when transported to other countries or cultures. Given the increasingly globalized nature of work, there have been repeated calls in the employee selection research literature to evaluate this question (e.g., Bartram, 2004; Caligiuri & Paul, 2010; Carey, Herst, & Chan, 2010; Herriott & Anderson, 1997; Lievens, 2008; Ryan & Ployhart, 2014; Schmitt, Cortina, Ingerick, & Wiechmann, 2003).

In the context of evaluating the effectiveness of a global selection instrument, two competing hypotheses examine if a selection instrument is valid in one country but not other countries (i.e., situational specificity) or the criterion-related validity observed in one country generalizes to other countries (i.e., validity generalization) (Salgado & Anderson, 2002). At present, research has primarily examined and found support for the cross-national generalizability of commonly used predictors such as cognitive ability (Salgado, Anderson, Moscoso, Bertua, & de Fruyt, 2003; Salgado, Anderson, Moscoso, Bertua, de Fruyt, & Rolland, 2003) and personality (Salgado, 1997; Salgado, 1998a) to the European Community. Since cognitive ability tests and personality assessments are two of the most commonly used selection instruments, outside of traditional hiring tools such as interviews and applications (Ryan et al., 1999), it is unsurprising that there is sufficient evidence to evaluate the cross-national generalizability of these predictors. It is, however, remarkable that little research has examined this question in countries outside of Europe or with other potentially useful personnel selection instruments.

One such tool is biographical data. Biographical data is a standardized self-report selection method that asks applicants questions about prior behaviors and experiences (Mael,

1991). It has the potential to be broadly useful in organizations for a number of reasons. First, previous research has found that biographical data inventories predict a variety of organizational outcomes, including performance ratings, objective performance, tenure, creativity, and training success (Bliesener, 1996; Schmidt & Hunter, 1998). Second, they have the potential to add incremental validity over interviews (Dalessio & Silverhart, 1994), cognitive ability tests (Mael & Hirsch, 1993), and Big Five assessments (Mount, Witt, & Barrick, 2000; Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004). This is important because a selection system's overall validity is a direct determinant of its practical value (Schmidt & Hunter, 1998). Lastly, biographical data inventories tend to have less adverse impact across racial and gender groups compared to other predictors such as GMA (Bobko & Roth, 2013; Bobko, Roth, & Potosky, 1999).

Presently, however, biographical data instruments are traditionally used less frequently by organizations because human resource (HR) professionals perceive these tools as less valid, practical, and legally defensible (Furnham, 2008; Ryan et al., 1999). This is unfortunate because research tends to suggest that these beliefs are unsubstantiated, at least within North America (Breugh, 2009; Mumford, Barret, & Hester, 2012; Schmitt & Golubovich, 2013). Therefore, the purpose of the present research is to address this gap and extend existing knowledge by evaluating the cross-national generalizability of biographical data inventories. Specifically, I comprehensively examine this question in two separate studies involving employees from 7 country clusters, across four continents and two different occupational groups (managers and manufacturing technicians), to determine if these instruments have utility in global high-stakes selection systems.

In the forthcoming sections, I discuss the strategic importance of personnel selection for multinational organizations and the unique challenges of designing and deploying global

selection systems. After, I examine available research on the cross-national specificity or generalizability of personnel selection instruments. Last, I provide a focused review of the biographical data literature and present evidence to support my hypothesis that carefully developed biographical data instruments will generalize to other countries. If supported, the findings would suggest that multinational organizations should consider including these instruments, along with other commonly used and validated predictors, when screening job applicants to improve the strategic value of their global selection systems.

The Strategic Importance of Personnel Selection

Personnel selection is a key function for human resources departments. It is concerned with identifying the individuals who will make up an organization's workforce. One important characteristic of an organization's personnel selection system is the extent to which it allows them to identify people who possess strategically important KSAOs. This is critical because the composition of KSAOs within an organization's workforce contributes to its success and likely serves as a source of competitive advantage. In the forthcoming section, I expand on how personnel selection influences an organization's human capital and how, in turn, it may become a source of competitive advantage for a firm.

Human resources management (HRM) is concerned with the effective management of people within an organization to achieve both individual and organizational objectives (Cascio & Aguinis, 2011). HRM includes a variety of functions related to hiring, training and developing, appraising, and retaining employees. When an organization designs and implements effective policies, practices, and systems to manage these HR functions, they have the potential to transform its human capital, or talent pool, into a source of sustained competitive advantage (Ployhart & Weekley, 2010). A firm can consider its human capital a sustained competitive

advantage to the extent that it is valuable, rare, inimitable, and non-substitutable (Wright, Dunford, & Snell, 2001; Wright, McMahan, & Williams, 1994). That is, a firm's human capital has strategic value to the extent that it is composed of employees who are valuable (i.e., help to achieve strategic organizational objectives) and have skills that are in limited supply and not easily duplicated or substituted.

An organization's human capital is created from the aggregation or emergence of individual employees' KSAOs to unit-level (e.g., firm, business unit, division, or group) competencies (Ployhart & Moliterno, 2011). Therefore, at the most fundamental level, a key determinant of whether or not a firm's human capital will emerge as a source of sustained competitive advantage is the extent to which individual employees have the right KSAOs to perform effectively on the job. These KSAOs include a variety of cognitive (e.g., GMA, skills, experience, job knowledge) and non-cognitive (e.g., personality traits, interests, values, motivation) constructs (Motowidlo, Borman, & Schmit, 1997). When employees' KSAOs are aggregated to higher levels of an organization, it may create valuable unit-level competencies. For example, an organization could screen and select sales agents based on Conscientiousness and Extraversion. Although these personality traits predict individual sales performance (Vinchur, Schippmann, Switzer, & Roth, 1998), they do not necessarily create competitive advantage for a firm. When combined with other HR practices (e.g., training, performance-based compensation), a strong organizational culture, and supportive management, these individual characteristics may emerge as a unit-level competency such as "customer-orientation" that emphasizes a willingness for all salespeople to go above and beyond to meet customers' needs (Ployhart & Weekley, 2010). Ultimately, the extent to which these unit-level competencies are valuable, rare, inimitable, and non-substitutable determines if an organization's human capital is

a source of sustained competitive advantage (Ployhart & Moliterno, 2011; Ployhart & Weekley, 2010).

When discussing the strategic value of HRM, it is important to distinguish between an organization's human capital and specific HR policies, practices, and systems because only the former are usually considered a potential source of competitive advantage (Wright et al., 1994; Ployhart & Weekley, 2010). HR practices are generally not sources of *sustained* competitive advantage because they are widely adopted, easily imitated, and substitutable. For example, if an organization decides to use a test of GMA to screen applicants, it is unlikely to be a long-term competitive advantage because GMA is widely known to be one of the single best predictors of job performance (Hunter & Hunter, 1984) and other organizations across the globe commonly use similar tests (Ryan et al., 1999). Despite this distinction, it is important to note that the overall validity of a selection system is one key determinant of its practical value to an organization (Schmidt & Hunter, 1998). Therefore, organizations seek to maximize the validity of a test battery with a variety of strategies (e.g., including multiple unique predictors), which may create a *short-term* advantage over competitors. Additionally, although HR practices are not intrinsically sources of sustained competitive advantage, they are vital to developing and maintaining an organization's talent pool and a necessary condition for ultimately transforming an organization's human capital into a high-performing workforce and an enduring competitive advantage (Combs, Liu, Hall, & Ketchen, 2006; Crook, Todd, Combs, Woehr, & Ketchen, 2011).

As previously mentioned, individuals' KSAOs are the foundational element in determining the strategic value of an organization's talent pool (Ployhart & Moliterno, 2011). Consequently, an organization that is interested in making its human capital a source of sustained

competitive advantage should invest in HR practices that ensure all employees have the KSAOs necessary to succeed on-the-job. The policies, practices, and systems that accomplish this generally fall within the HR functions of personnel selection and training and development (Cascio & Aguinis, 2011).

Personnel selection is primarily concerned with ensuring that an organization brings in people who have the right KSAOs. Information about candidates' standing on a variety of job-relevant KSAOs (e.g., GMA, interpersonal skills, Conscientiousness, psychomotor ability) is elicited via one or more selection methods (e.g., paper-and-pencil test, biographical data inventory, assessment center, situational judgment test, and interview) (Arthur & Villado, 2008). As part of a selection system, an organization may include a variety of methods to obtain information about candidates on multiple job-relevant constructs. Then, based on a candidate's scores across the various methods, the organization will decide to reject or extend a job offer.

In contrast, training and development is focused on increasing the effectiveness of current employees by providing them opportunities to learn new job-relevant competencies (i.e., clusters of KSAOs) so they are equipped to handle the demands of current and future jobs (Cascio & Aguinis, 2011). Depending on the training objectives, an organization can use a variety of different programs tailored to the specific audience, such as new hires, supervisors, executives, teams, or international assignees (Salas, Tannenbaum, Kraiger, & Smith-Jentsch, 2012).

In either case, an organization first determines, via work analysis, what KSAOs are most important to succeed on-the-job. Then, they design an appropriate strategy to help ensure that all employees are able to contribute to the organization's effectiveness. All else being equal (e.g., base rate, selection ratio, variability in job performance), there is a trade-off between these two HR functions. Specifically, organizations must decide how to allocate resources between

selection and training and what job-relevant KSAOs they want to select candidates on versus train new hires on to enable a high-performing workforce. For example, an organization could have a more stringent selection process and hire people who already possess the necessary KSAOs and invest fewer resources to training and development. Alternatively, an organization could focus on hiring people that only meet the minimum job qualifications and devote more resources to training them once they are onboard (i.e., build from within). If a hypothetical organization is given the choice between selection and training, Cascio and Aguinis (2011) note that “the best strategy is to choose selection” (p. 46). They argue that selecting individuals who are probabilistically more likely to succeed on the job from the beginning will also be able to learn new information quickly in future training compared to lower quality candidates¹. Therefore, if an organization is interested in transforming their human capital into a source of sustained competitive advantage, it is important that they invest in a high-quality selection system, aligned with their strategic plan, so all new hires have the necessary KSAOs to execute the firm’s strategy. Based on this rationale, personnel selection is the focus of this study because it is the first opportunity for an organization to influence the composition of KSAOs among its employees.

The preceding discussion focused on the strategic importance of personnel selection in organizations. Specifically, an organization’s human capital is created by aggregating individual employees’ KSAOs to higher levels of the organization. While HR functions, policies, and practices are generally not considered sources of sustained competitive advantage, they influence the composition of KSAOs within an organization’s talent pool. The two functions with the largest impact on this are personnel selection and training. The present study is focused on

¹ This is not meant to minimize the importance of training and development. Rather, both functions have a critical place in an organization’s overall HRM strategy.

personnel selection because it is an organization's first opportunity to influence employees' KSAOs. It is critical for an organization to invest in a high-quality system that allows them to hire employees with the KSAOs needed to realize their operational strategy. The process of developing and deploying such a system is challenging for all organizations but becomes more so when an organization operates across geographical boundaries (Caligiuri & Paul, 2010; Ryan & Tippins, 2009). There are a number of benefits, however, if an organization can deploy a global system. In the next section, I discuss global selection strategies in multinational organizations and associated challenges and benefits.

Personnel Selection in Multinational Organizations

Multinational organizations face unique challenges when developing and deploying global selection systems. Specifically, they differ from domestic firms in that they operate across geographical borders and their employees, clients, and vendors come from diverse cultural backgrounds. In the context of HR, this requires multinational organizations to balance the need for centralized operations and local responsiveness (Caligiuri & Paul, 2010). In this section, I discuss personnel selection strategies and work analysis approaches for multinational organizations. Then, I discuss the unique challenges faced by multinational organizations when coordinating hiring activities across geographical and cultural boundaries, including relevant cross-cultural research. Lastly, I discuss the potential benefits multinational organizations may realize by taking a global or standardized approach to personnel selection.

Multinational personnel selection strategies and work analysis.

As noted by Caligiuri and Paul (2010), there are three general strategies a multinational organization can pursue when designing a global selection system. These strategies differ in how they balance centralization, or the extent to which a global team controls the selection system,

and decentralization, or the extent to which the system varies across different regions. First, an organization may elect to pursue a decentralized or multidomestic strategy, in which the organization customizes solutions to each local market. Second, a centralized or global strategy focuses on integration and consistency across operating regions. Last, a transnational strategy balances these two competing demands. Relevant to the current research, the cross-national generalizability of a selection instrument is most relevant for an organization utilizing a strategy that includes or emphasizes centralization. In this section, I discuss these multinational selection strategies in more detail.

A multidomestic strategy emphasizes local responsiveness over integration. This is primarily due to the organization's structure, which is decentralized. Specifically, regional offices tend to resemble a local firm and there is little input from global headquarters. The benefit of this approach is that a multinational organization is able to compete locally based on the unique demands of each country or region. In terms of the personnel selection system, an organization pursuing a multidomestic strategy is likely to use different selection tools and systems across countries based on unique local demands. That is, the subsidiary office, rather than global headquarters, is likely to make decisions regarding the constructs and methods used to hire employees. This approach is likely to be more expensive because organizations cannot achieve economies of scale with their selection tools and makes it difficult for a multinational organization to maintain consistency around the world (Caligiuri & Paul, 2010). Additionally, it makes it difficult to compare candidates across countries and select employees for international assignments within the firm.

Second, a global strategy emphasizes centralization and integration at the expense of decentralization and local responsiveness. Typically, an organization pursuing a global selection

strategy has a central headquarters that manages worldwide business units. The benefit of this approach, from an operational standpoint, is that it allows organizations to deliver products efficiently and consistently around the world. With this strategy, headquarters sets global standards, and country or regional branches adhere to these standards. This helps to ensure a consistent experience for clients or customers around the world, which ultimately serves as a competitive advantage for a global organization. For example, many people visit McDonald's when traveling because they know they will get a similar experience regardless of location. For personnel selection, the emphasis on uniformity likely means that employees and new hires will need to possess a similar or the same profile of KSAOs across different operating regions. In practice, this means that a multinational organization may have a "centrally developed selection system" involving the same battery of selection tools to ensure that all candidates, globally, have the necessary KSAOs to contribute to the organization's effectiveness (Caligiuri & Paul, 2010, p. 782).

Last, a transnational selection strategy attempts to balance global integration and local responsiveness by establishing an "interdependent global network of subsidiaries" that work closely with headquarters to make operational decisions that meet both global and local business needs (Caligiuri & Paul, 2010, p. 784). Transnational personnel selection systems pursue consistency for strategic reasons, while allowing local HR offices to make minor modifications to meet their specific needs. For example, an organization may deploy one standardized selection tool across all regions but permit each subsidiary to modify when (e.g., before or after interviews) and how (e.g., unproctored or proctored) the tool is used based on unique cultural demands. This approach helps to ensure that globally mandated selection tools are useful and accepted in each region, while also realizing the benefits of standardization.

Regardless of the selection strategy a multinational organization pursues, the first step in designing a useful system is to determine the profile of KSAOs important for new hires to possess in a given job. Organizations use a variety of work analysis methods to answer this question. These techniques can be grouped into two broad categories—job analysis and competency modeling (Brannick, Levine, & Morgeson 2007).

Both approaches are similar in that they seek to understand the critical KSAOs needed to succeed in a job. One point of differentiation, however, is that the typical competency modeling approach identifies broad clusters of KSAOs (i.e., competencies) that are strategically important to achieve an organization's goals. In contrast, job analysis identifies the KSAOs required to perform a specific job without reference to an organization's strategy (Sanchez & Levine, 2009; Schippmann et al., 2000). That is, competency modeling focuses on identifying KSAOs aligned with the execution of an organization's strategy and job analysis primarily focuses on describing job-specific KSAOs without reference to an organization's strategic goals. In the context of transforming an organization's human capital into a source of sustained competitive advantage, competency modeling is likely to be more beneficial because an organization screens and selects job applicants on KSAOs linked to its strategy.

A second difference between competency modeling and job analysis is that competency modeling assumes there are behavioral dimensions common across multiple jobs in an organization and job analysis focuses on describing the KSAO required for one specific job (Sanchez & Levine, 2009). Using the language of personnel selection strategy, competency modeling typically results in one centralized KSAO profile across a large group of job families and job analysis produces decentralized KSAO profiles for each job or job family within an organization. The benefit of a standardized approach is that an organization can easily

communicate its employment brand and what behaviors it strategically values to all candidates, new hires, and current employees. As an example, FedEx has a competency called “discretionary effort”. This competency encourages employees to problem solve and generate novel ideas that delight customers. This is consistent with their business strategy to exceed customer expectations (Sanchez & Levine, 2009). Ultimately, the purpose of competency modeling is to communicate the organization’s culture and provide employees guidance on how they can contribute to the broader business goals with their daily behavior.

While these differences appear to suggest that organizations should favor competency modeling, it is important to point out that the typical competency modeling project is less rigorous from a research perspective (Schippmann et al., 2000). Therefore, if an organization elects to use competency modeling, the quality of the end product (i.e., the final competency model) is likely to be improved if researchers incorporate elements of job analysis such as obtaining feedback from subject matter experts (SMEs) and providing SMEs information about the tasks involved in each job (Lievens, Sanchez, & De Corte, 2004). Ultimately, whether an organization elects to use job analysis or competency modeling, the result is a list of KSAOs that are important for job performance. An organization then uses this information to develop a selection system, involving one or more methods, that reliably and accurately captures candidates’ standing on the identified KSAOs to predict who is likely to succeed on the job (Binning & Barrett, 1989; Cascio & Aguinis, 2011).

When an organization elects to use competency modeling as the foundation for its selection system, they are saying that all new hires must possess a specific profile of strategically important competencies (in addition to job-specific technical dimensions). Therefore, when relating this discussion to the broader context of a multinational organization’s selection strategy,

competency modeling is aligned with either a global or a transnational approach since it is focused on centralization. In these situations, a critical consideration for multinational organizations is the extent to which a selection tool or system is equally effective across operating regions. If an organization can design and deploy an instrument that generalizes across countries, they stand to benefit in a number of ways compared to a more decentralized approach—points which will be expanded upon in the next section.

It is important to emphasize that it is uniquely challenging for multinational organizations to develop and deploy a centralized selection system compared to a domestic organization. When a company that operates in one country utilizes a standardized approach, it is reasonably easy to develop an effective selection system because there is strong evidence that the validity of similar assessments generalizes to other settings (e.g., Schmidt & Hunter, 1977; Schmidt & Hunter, 1998). However, this approach becomes more difficult for a multinational organization because the test(s) administered as part of the selection process may not be equally effective when administered in other regions due to country, culture, or language differences (Lievens, 2008; Ryan & Tippins, 2009; Ployhart & Weekly, 2010). Consequently, it is important for a multinational organization to evaluate if their selection system is equally effective in other cultures or geographical regions. If they can design a system that has generalized validity across operating regions, they stand to benefit in a number of ways.

In sum, I discussed three strategies multinational organizations can pursue—multidomestic, global, and transnational—when designing their selection systems. A critical element of this process is to determine what KSAOs are important for job performance. Job analysis and competency modeling are the two primary ways organizations answer this question. In contrast to job analysis, competency modeling focuses on identifying the strategically

important KSAOs common to many jobs throughout an organization. When integrating selection strategies and work analysis methods, I argued that global and transnational strategies are aligned with competency modeling and these approaches are more likely to aid an organization seeking to transform its human capital into a source of competitive advantage. In the next section, I expand on the aforementioned unique challenges and opportunities for multinational organizations when developing and deploying a centralized selection system.

The challenges of global selection systems.

There are many unique challenges when developing and implementing global selection systems. Rather than provide an exhaustive review of all possible methodological, measurement, and cultural issues, the purpose of this section is to highlight elements that are most relevant for organizations when deploying centralized selection instruments and evaluating their cross-national generalizability. For more detailed information, see Ryan and Tippins (2009), Schaffer and Riordan (2003), or Spector, Liu, and Sanchez (2015).

Foremost among the challenges, it is necessary to determine if there is a common job across countries or cultures (Ryan & Tippins, 2009). Specifically, an organization must determine, via job analysis or competency modeling, if the tasks involved in a given job and the KSAOs required to perform those tasks are similar across locations. An organization can gather data to evaluate this question in a variety of ways including observations, focus groups, or questionnaires (Brannick et al., 2007). After, it is important to analyze the results across countries to ensure commonality. If a job is similar, then it is appropriate to use the same selection system globally. However, if there are large differences in the KSAO requirements for the same job across regions, applying the same selection instrument would be less effective

(Lievens, 2008). In this situation, it would be more appropriate for an organization to design a separate selection system focused on the unique KSAO requirements of the job in each country.

A second challenge when deploying a global selection instrument concerns using an assessment method that is cross-culturally acceptable to candidates (Caligiuri & Paul, 2010). This is important because reactions to pre-employment assessments have been linked to candidates' intentions to accept job offers, recommend the organization to others, and file legal complaints (Hausknecht, Day, & Thomas, 2004). Fortunately, a recent meta-analysis suggests that applicants' reactions to selection instruments generalizes across countries (Anderson, Salgado, & Hülshager, 2010). That is, candidates from different countries tend to have similar impressions of various selection instruments, with some instruments being most preferred (work samples, interviews), favorably evaluated (resumes, cognitive tests, references, biodata, personality), and least preferred (integrity tests, personal contacts, graphology).

Third, when designing a global selection system, it is important to consider logistical constraints in some regions (Caligiuri & Paul, 2010). For example, when conducting proctored online testing, some locations may not have access to testing rooms with multiple computers connected to the internet. Alternatively, if using unproctored testing, participants may not have access to computers and may only be able to access the selection instrument on a mobile device. In these situations, it may be necessary to modify the assessment protocol to accommodate limitations in testing sites or make unproctored assessments mobile responsive. Another common logistical constraint when administering pre-employment assessments is differences in internet speed between countries. This may cause a test to lag, which could be problematic for adaptive and/or timed tests, and require additional investment in a web acceleration tool (Akamai, 2014).

After considering these issues related to test design and deployment, it is important to ensure that the constructs assessed by a selection instrument are equivalent, or invariant, across countries or cultures (Vandenberg & Lance, 2000). This step is focused on ensuring that each culture's unique norms, values, attitudes, and experiences do not influence the measurement of the construct(s) of interest. At a general level, this involves two elements—semantic equivalence and measurement invariance (Schaffer & Riordan, 2003).

When applicable, semantic equivalence focuses on ensuring a measure is accurately translated into a new language. Best practices recommend the use of blind back-translation (Schaffer & Riordan, 2003). This involves translating a measure into the new language. Then, a second person, who has not seen the original version, translates the newly translated measure back into the original language. Last, the back-translated version of the measure is compared to the original and discrepancies among the versions are discussed and resolved. Before translating a measure into a new language, however, it is important to review items to ensure they are relevant across the countries in question and do not include idioms, phrases, or terms unique to one culture (Mumford, Costanza, Connelly, & Johnson, 1996).

Measure invariance expands on the concept of equivalence by examining the extent to which a test measures the same construct with the same precision across groups (Vandenberg & Lance, 2000). That is, it examines the equivalence of a measure to see if the construct(s) of interest and response options have a similar meaning across cultures. It is most commonly tested via multiple-group confirmatory factor analysis (MG-CFA) or item response theory (IRT). In psychology, MG-CFA is used more frequently (Spector et al., 2015). This procedure involves testing a series of nested models to examine the between-group equivalence of the variance-covariance matrices, factor structure (i.e., configural invariance), factor loadings (i.e., metric

invariance), intercepts (i.e., scalar invariance), and uniquenesses (Millsap & Everson, 1993; Vandenberg & Lance, 2000). It is important to note that the more stringent tests (e.g., equal uniquenesses) are very strict and it is often unrealistic for these types of invariance to hold in practice “except in highly contrived situations” (Chan, 2008, p. 68).

Relevant to the current research, in a typical biographical data inventory, questions and response options are multidimensional because they acknowledge that behavior is complex and determined by multiple constructs. Moreover, test items commonly use a variety of dichotomous, multiple-choice, and Likert-type response formats (Schmitt & Golubovich, 2013). From a measurement perspective, this makes it problematic to formally test measurement invariance because items are likely to cross-load on multiple factors and the resulting mix of nominal, ordinal, and interval response options violates the assumption of multivariate normality. Given these constraints, development of a biographical data inventory for global use should be especially concerned with conceptual equivalence of constructs in terms of similar meanings across countries (Lievens et al., 2015). For example, in Japan, a construct like “customer service orientation” may emphasize the importance of anticipating the needs of customers. In contrast, in Italy, the same construct may emphasize providing candid feedback to customers because they expect a representative to share his or her expertise. However, in both countries, there is a common desire for a representative to work hard to meet the customer's needs and resolve a problem. Therefore, if a company wants to use a global biographical data instrument for this job, they should develop a measure around common ideas underlying the construct to ensure conceptual equivalence (Caligiuri & Paul, 2010). Furthermore, since the purpose of the present research is to examine the equality in direction and magnitude of the relationship between scores on a biographical data inventory and job performance, rather than mean differences on the

predictor between countries or to identify between-country differences, an inability to evaluate measurement invariance is somewhat less problematic (Chan, 2008; Lievens et al., 2015).

The last challenge to deploying a global selection system is related to between-country differences in the laws that influence personnel selection practices and the storage of candidate data gathered during the hiring process. Managing this complexity and ensuring compliance is challenging for a variety of reasons. For example, while nearly all countries protect applicants from discrimination on the basis of race, color, religion, gender, national origin, age, and disability status, many countries have passed laws prohibiting discrimination against additional groups (e.g., HIV status in South Africa; caste in India; philosophical beliefs in Turkey) (Sackett et al., 2010). Therefore, it is important to ensure that selection instruments do not discriminate against any protected groups. As another example, countries differ in the evidence required to dispute a discrimination claim. In some countries (e.g., Chile, Italy), the evidentiary standard is unclear and it is generally not necessary to provide empirical evidence to support the use of a selection instrument. In contrast, in the United States (U.S.), there is clear guidance that an organization must demonstrate that a selection practice is job-related (e.g., exhibits criterion-related validity) and consistent with business necessity. Lastly, data privacy regulations and laws vary at the country level and may create challenges for global companies when storing candidate assessment data. For example, Russia requires personal data of Russian citizens to be stored on servers located in Russia. Ultimately, these examples underscore the complexity of personnel selection for multinational organizations and the importance of understanding the practical implications of differences between operating regions. Collectively, all of these issues highlight the importance of culture in the context of personnel selection and the need for multinational organizations to be aware of how cultural values may influence workplace behavior.

The importance of culture for multinational personnel selection.

Up to this point, I have referenced both country and culture when discussing the generalizability of a selection instrument. It is important, however, to clarify the distinction between these two terms. One common definition of culture is the “shared motives, values, beliefs, identities, and interpretations or meanings of significant events that result from common experiences of collectives that are transmitted across generations” (House et al., 2004, p. 15). In contrast, country emphasizes the geographical boundaries of a political entity.

Country is frequently used as a proxy for culture in cross-cultural research because there is an assumption that people prefer to interact with other individuals and governmental institutions consistent with their own values and beliefs (Peterson & Smith, 1997; Schaffer & Riordan, 2003). While this practice is convenient and may be suitable for multinational organizations, it is important to note that country and culture are not equivalent for a variety of reasons. Examples include national boundaries drawn by powerful outside parties without reference to shared culture (e.g., parts of Africa), multiple subcultures within one country (e.g., Canada, Switzerland), or political differences within a country (e.g., the U.S. before the Civil War) (Peterson & Smith, 1997). Despite this distinction, from a practical standpoint, most multinational organizations operate along country or regional boundaries. Therefore, investigating the cross-national (rather than the cross-cultural) generalizability of a selection instrument is informative from an applied perspective but may make it difficult to explain observed differences when validity generalization is not supported. This, however, does not negate the importance of cultural values. Since most of this research is at the national-level, it is useful for organizations to understand how these cultural values influence differences in work behavior between countries. Accordingly, I briefly review the three main cultural frameworks.

Hofstede (1980, 2001) conducted research on differences in cultural values within subsidiaries of one multinational organization (IBM) across 72 countries. His research revealed five dimensions. *Power Distance* is the extent to which less powerful members of society accept that power is distributed unequally. *Uncertainty Avoidance* is the degree to which people in a society feel uncomfortable with ambiguity. Countries low in Uncertainty Avoidance have a relaxed attitude and countries high in Uncertainty Avoidance emphasize adherence to social norms and bureaucratic practices. *Individualism-Collectivism* is a continuum indicating the degree to which individuals in a country prefer to integrate into social groups. At the high end, people are only expected to take care of themselves and their immediate families. In contrast, countries high in collectivism tend to take care of other in-group members. *Masculinity-Femininity* refers to a society's preference for competition versus cooperation. Societies high in Masculinity emphasize achievement, assertiveness, and material rewards for success. In contrast, Femininity emphasizes cooperation, modesty, and quality of life. Last, *Long- versus Short-Term Orientation* refers to how a society maintains links with its past. Societies with a Short-Term Orientation maintain traditions and norms and those with a Long-Term Orientation encourage thrift and perseverance, and look toward the future.

Second, Schwartz (1992, 1994) conducted research on 41 cultural groups in 38 countries. His research uncovered five societal values. *Conservatism or Embeddedness* reflects the extent to which a country is primarily concerned with security, conformity, and tradition. *Intellectual Autonomy* refers to the extent that a society emphasizes the independent pursuit of ideas and intellectual goals. *Affective Autonomy* refers to the extent that a society has an affective emphasis on stimulation and hedonism. *Hierarchy* refers to the tendency for a country to accept an unequal distribution of power and resources; in contrast, *Egalitarianism* emphasizes voluntary

commitment to social justice. These dimensions are similar to the high and low poles, respectively, of Power Distance. Last, *Mastery* focuses on getting ahead through self-assertion and *Harmony* emphasizes fitting into the environment. These dimensions are similar to In-Group Collectivism (House et al., 2004) and Individualism-Collectivism (Hofstede, 1980, 2001).

Finally, the Global Leadership and Organizational Behavior Effectiveness (GLOBE) project developed the most comprehensive model of cultural values as part of a large international study on culture and leadership (House et al., 2004). A team of 180 researchers gathered data from over 800 organizations in 62 countries to examine societal values. Results indicated nine cultural dimensions, with some similarities to Hofstede's (1980, 2001) framework. *Power Distance* and *Uncertainty Avoidance* parallel Hofstede's definitions. While Hofstede identified one dimension of Individualism-Collectivism, House and colleagues (2004) identified two dimensions: *Institutional Collectivism*, which refers to the extent to which a society emphasizes collective distribution of resources and collective action, and *In-Group Collectivism*, which is similar to Individualism-Collectivism. *Gender Egalitarianism* is similar to Hofstede's Masculinity-Femininity. In addition and separate from Gender Egalitarianism, *Assertiveness* refers to the degree that individuals are assertive, aggressive, and confrontational when interacting with others. *Future Orientation* is similar to Hofstede's Long- versus Short-Term Orientation. Lastly, House and colleagues identified two unique cultural dimensions: *Performance Orientation*, or the extent to which a society encourages innovation, achievement, and performance improvement, and *Humane Orientation*, which focuses on the extent to which a society encourages individuals to be fair, friendly, generous, and kind to others.

In addition to identifying these nine cultural dimensions, the researchers reported the mean score for each country on the nine cultural values and grouped culturally similar countries

together, using discriminant analysis, into 10 clusters: Anglo, Latin Europe, Nordic Europe, Germanic Europe, Eastern Europe, Latin America, Middle East, Sub-Saharan Africa, Southern Asia, and Confucian Asia. These country clusters are especially relevant to the present research, as they will be used to group culturally similar countries together when evaluating the cross-national generalizability of the two biographical data inventories.

It is important for multinational organizations to be aware of these cultural values because they have the potential to influence candidates' behavior during the selection process. For example, candidates high in Individualism may be more likely to talk about their individual accomplishments during an interview and those high in Collectivism might focus on group accomplishments (Ryan & Tippins, 2009). Research has also found that these cultural values have practical importance. Meta-analytic work shows that they relate to a variety of individual-level (e.g., contextual performance, innovation, entrepreneurial behavior) and country-level (e.g., wealth, innovation, economic growth) outcomes (Taras, Kirkman, & Steel, 2010). In addition, in the context of personnel selection, research suggests that cultures high in Uncertainty Avoidance tend to use tests and interviews more frequently (Ryan et al., 1999). Lastly, research shows that there are some differences in response styles between cultures. Specifically, individuals from East Asian cultures are more likely to use the mid-point of Likert-scales compared to individuals from North America (Chen, Lee, & Stevenson, 1995) and individuals from Hispanic countries are more likely to endorse the endpoints of a response scale (Hui & Triandis, 1989). Ultimately, the purpose in briefly reviewing these three cultural values frameworks is to highlight meaningful cultural and national differences and the need for multinational organizations to be aware that these differences may influence the acceptability, utility, and adoption of various selection instruments.

This section highlights the complexity of global personnel selection. Multinational organizations must be aware of these methodological, logistic, and legal issues and the potential for cultural values to influence candidate and employee behavior throughout the selection process. If organizations can address these challenges, however, they have the potential to realize a number of benefits by designing and deploying a global selection system.

The benefits of global selection systems.

While the aforementioned challenges may appear insurmountable, research suggests that it is possible for organizations to design and deploy centralized selection systems (Lievens, 2008). When multinational organizations achieve this objective, there are many strategic and practical benefits. The purpose of this section is to highlight a few of the ones most pertinent to the current research.

From a strategic perspective, global selection systems have the potential to help multinational organizations transform their human capital into a sustained competitive advantage for a number of reasons. A standardized selection system helps to ensure that all new hires possess strategically important KSAOs. This allows organizations to provide standardized products and services to customers around the world. As previously mentioned, this is frequently a source of competitive advantage for global organizations. Second, a global selection system allows multinational organizations to identify new hires who are able to succeed in different geographical regions within the company (Carey, Herst, & Chan, 2010). This is important for organizations when evaluating the capabilities of their workforce and anticipating future hiring needs. It also makes it easier to identify internal candidates for international assignments or promotions. Third, a global selection system helps to provide a consistent image to candidates about the organization worldwide. This enables them to provide information about their

employment brand early in the hiring process so they can attract high-quality candidates that fit with their organization (Ryan & Tippins, 2009). Lastly, a global system ensures all candidates are treated equally, thereby increasing procedural justice perceptions, which are associated with increased organizational attractiveness and offer acceptance intentions (Hausknecht et al., 2004).

There are also a number of practical advantages to using a global selection system. First, it reduces development and administration costs (Ryan & Tippins, 2009). By using a standardized assessment process, an organization saves money on the research and validation of new selection instruments in each region. In addition, administration costs decrease when there is a single selection system because online hosting costs tend to decrease as test volume increases. Ultimately, the implication of these cost savings is that it makes the hiring process more efficient by requiring fewer resources (e.g., one test platform, reduced administrator training, fewer integrations with the applicant tracking system) and helps reduce the time to hire a candidate—a key operational metric for recruiting professionals. Despite these benefits, it is important to acknowledge that a global selection strategy is not always appropriate for multinational organizations and they should pursue a strategy that fits with the realities of their operating environment (Caligiuri & Paul, 2010).

In summary, in this section I first discussed three personnel selection strategies for multinational organizations—multidomestic, global, and transnational—and how they differentially weight the competing demands of centralization and decentralization. After, I discussed the importance of identifying strategically important KSAOs, via either job analysis or competency modeling. Last, I reviewed a variety of methodological, logistic, and legal challenges and the benefits of deploying a global selection system. One important open question is the extent to which it is practical to expect a selection system to be equally effective when

deployed in other countries. In the next section, I discuss this question at a conceptual level and review relevant research.

The Cross-National Generalizability of Personnel Selection Instruments

In an increasingly flatter world, organizations are frequently operating beyond national borders (Friedman, 2007). International partnerships, joint ventures, mergers, and acquisitions have expanded the labor market from a domestic perspective to an international one. In this environment, multinational organizations demand personnel selection tools that can be effective across different countries, geographical regions, and cultures (Lievens, 2008). In this section, I discuss the concept of validity generalization and review research on the cross-national generalizability of selection tools.

At a conceptual level, a selection instrument will be practically useful (i.e., demonstrate criterion-related validity) to the extent that there is overlap between the predictor and criterion construct (Binning & Barrett, 1989). Said differently, validity is focused on matching predictor behaviors with the performance domain. For example, a personality instrument that assesses applicants' standing on the Big Five personality trait of Conscientiousness will demonstrate validity to the extent that behaviors associated with being high on Conscientiousness (e.g., showing up on time, working hard, planning) are also associated with higher levels of job performance.

In an international context, it is important to consider if the constructs assessed by a predictor instrument are aligned with the definition of performance within each culture. As an example, Lievens, Harris, Van Keer, and Bisqueret (2003) examined the validity of two assessment center exercises among a sample of European executives who were picked to work in Japan. Results indicated that a leaderless group discussion but not a presentation exercise

predicted subsequent job performance, even though both exercises assessed the same dimensions. When explaining this finding, Lievens et al. (2003) noted that the leaderless group discussion was more characteristic of the team-based Japanese culture and the exercise provided numerous opportunities to evaluate how a candidate would contribute to a team in Japan. In contrast, the presentation provided fewer opportunities to evaluate candidates in an environment that matched the performance domain (as cited in Lievens, 2008). Ultimately, this highlights the importance of understanding the criterion when evaluating if a predictor instrument is likely to generalize to a new country. In addition, it is important that the behaviors assessed by the predictor instrument are relevant within each country. Ultimately, this highlights the importance of conducting cultural reviews to ensure items tap constructs and behaviors relevant across the countries in question (Mumford, Costanza, Connelly, & Johnson, 1996).

At a more theoretical level, using a selection instrument in a different culture is similar to using a selection instrument in any new context that is different from the one in which it was originally developed and validated. From the 1920s to the mid-1970s, there was a widely held belief that a test's validity did not generalize to new jobs, organizations, or occupations because of subtle differences between jobs. Counter to this assumption, Schmidt and Hunter (1977) proposed that the observed variability in the validity of selection instruments was primarily due to statistical and measurement artifacts (e.g., criterion unreliability, range restriction, sampling error), rather than meaningful differences in the predictor or criterion. More formally, in the context of evaluating the cross-national generalizability of a selection instrument, researchers have proposed two competing hypotheses. First, the situational specificity hypothesis states that there is variability in validity coefficients between countries and these differences persist after correcting for statistical artifacts. In practice, this could mean that a selection instrument exhibits

significant criterion-related validity in one country but not in another or the magnitude of the predictor-criterion relationship is significantly different between two or more countries. In contrast, the validity generalization hypothesis says that differences in observed validities between countries are due to statistical artifacts. Once coefficients are corrected for these artifacts, the validity of a selection instrument in one country will generalize to another country (Schmidt & Hunter, 1977).

Using psychometric meta-analysis, researchers have aggregated the findings of individual validation studies to evaluate these two competing hypotheses. Results indicate that, after accounting for common statistical artifacts such as measurement error in the criterion, sampling error, and range restriction (when using incumbent samples), the variability in the validity of a given selection instrument tends to be small for the same job in different settings and across different jobs (Hunter & Hunter, 1984; Schmidt & Hunter, 1998). It is important to note, however, these findings have been “unreservedly cited by personnel psychologists in other countries and appear to have been unquestioningly accepted as being generalizable to different national contexts” (Herriot & Anderson, 1997, pp.27-28). This is problematic given the potential for cultural and legislative differences to influence the validity of selection instruments. Therefore, I review research that has evaluated the cross-national generalizability of personnel selection instruments to identify knowledge gaps.

Cognitive ability and personality.

Globally, cognitive ability and personality are two of the most commonly used assessment tools (Ryan et al., 1999). It makes sense that most research on the cross-national generalizability of selection instruments focuses on these tools. The most comprehensive research on this topic has been from Salgado and his colleagues. Regarding cognitive ability,

they examined the extent to which the validity of general mental ability measures generalized to the European community, including Belgium, France, Germany, Ireland, the Netherlands, Spain, and the United Kingdom (Salgado, Anderson, Moscoso, Bertua, & De Fruyt, 2003; Salgado, Anderson, Moscoso, Bertua, De Fruyt, & Rolland, 2003). The researchers found that GMA had generalized validity across a variety of occupational groups and European countries when predicting both training success and job performance. Compared to U.S. meta-analyses (e.g., Hunter & Hunter, 1984; Schmidt & Hunter, 1998), the corrected validity coefficients for Salgado and colleagues' results were similar (.56 vs. .54) or slightly higher (.62 vs. .51) than findings in the United States when predicting training success and job performance, respectively. Salgado et al. (2003) also replicated the finding that job complexity moderates the GMA-job performance relationship such that the validity is higher for more complex jobs (Hunter & Hunter, 1984). More recently, Oh (2009) examined the cross-national generalizability of cognitive ability to a new region, East Asia, which spanned five countries: South Korea, Japan, China, Taiwan, and Singapore. He found that, consistent with previous research in the United States (e.g., Hunter & Hunter, 1984) and Europe (Salgado et al., 2003), the validity of cognitive ability measures generalized across East Asia, the magnitude of the relationship was fairly similar across countries, and the relationship was moderated by job complexity. Collectively, this evidence supports the validity generalization hypothesis for cognitive ability instruments to these regions.

Regarding personality, Salgado (1997) examined the cross-national generalizability of the Big Five personality dimensions to predict job performance in European samples. Using meta-analysis, he found that Conscientiousness and Emotional Stability were valid predictors across performance criteria (i.e., job performance rating, training proficiency, and personnel record) and occupational groups (i.e., professionals, police, managers, sales, and skilled labor). Additionally,

the Big Five traits of Openness to Experience and Agreeableness were significant predictors of training performance. Lastly, some Big Five traits were predictors of performance for specific occupations. For example, Extraversion was a valid predictor for managers and police, Openness to Experience was a valid predictor for police and skilled labor, and Agreeableness was a valid predictor for professionals, skilled labor, and managers. These results are generally consistent with North American meta-analyses of the Big Five (Barrick & Mount, 1991; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Tett, Jackson, & Rothstein, 1991) and nearly identical to a second-order meta-analysis of the Big Five (Barrick, Mount, & Judge, 2001). In a subsequent international meta-analysis, Salgado (1998a) expanded on these findings and found, consistent with results in North America (e.g., Schmidt & Hunter, 1998), Conscientiousness and Emotional Stability explained incremental variance over general mental ability when predicting performance in the European community. Similarly, Oh (2009) examined the validity generalization of the Big Five to East Asia. He found that Extraversion and Conscientiousness, in that order, had the strongest validity. This is in contrast to Salgado (1997; 1998a), who found that similar to the United States (e.g., Barrick et al., 2001), Conscientiousness and Emotional Stability were most robust across criteria and occupations. When interpreting this difference, Oh (2009) noted that East Asian cultures have a stronger emphasis on interpersonal relationships in the workplace compared to North America and Europe. This discrepant finding emphasizes the importance of evaluating the validity of a selection instrument in each country or region in which it will be deployed. In combination, at present, there is evidence to support the cross-national generalizability of cognitive ability tests and some of the Big Five personality traits to a variety of European and East Asian countries.

Person-organization fit.

Person-organization fit refers to the compatibility between individuals and their work environments (Kristof, 1996). Oh and colleagues (2014) examined the validity of four types of person-organization fit in North American, Europe, and East Asia. They found that relationships between person-organization and person-job fit (i.e., rational fit) with a variety of outcomes (e.g., intent to quit) were relatively stronger in North American and Europe compared to East Asia. In addition, relationships between person-group and person-supervisor fit (i.e., relational fit) and a variety of outcomes were relatively stronger in East Asia compared to North America. The authors suggest that these differences may have occurred because relational (rational) fit is more important in collectivist (individualistic) and high power distance (low power distance) cultures.

When predicting job performance, the authors found that person-environment fit measures had modest but significant positive relationships in East Asia and North America. Unfortunately, due to a lack of primary studies, they were unable to examine the relationship between person-environment fit and job performance in Europe. Cross-regional comparisons indicated the magnitude of the relationship for rational fit measures was not significantly different between East Asia and North America but person-job fit was not significantly related to job performance in North America. When comparing results for measures of relational fit, results indicated that the both types significantly predicted job performance in both region but the magnitude of the correlation was significantly stronger in East Asia for person-group fit.

Lastly, when examining the relative importance of each type of fit for predicting job performance, results indicated that relational fit was about 1.5 times more important in North America (Relative Weight = 74%) compared to East Asia (Relative Weight = 49%). When looking at specific types of person-environment fit, person-supervisor fit (Relative Weight =

59%) was most important in North America and person-job fit (Relative Weight = 42%) was most important in East Asia. Despite these differences, the total variance explained ($R^2 = .12$) was the same for each region. This suggests that if one were to assess all four types of person-environment fit in a selection setting, the validity of the instrument would generalize across North America and East Asia. Collectively, these findings suggest there are some differences in P-O fit measures between North America and East Asia but the overall validity is similar for each region.

Situational judgment tests.

Up to this point, research on the cross-national generalizability of selection instruments has focused on predictors that capture signs of behavior rather than those that capture samples of behavior (Wernimont & Campbell, 1968). Sign-based predictors (e.g., cognitive ability test, personality inventory) capture candidates' standing on constructs that are expected to related to the criterion of interest and provide an indication of future performance. In contrast, sample-based predictors (e.g., biographical data, assessment center) present job-relevant scenarios to candidates and provide opportunities for candidates to demonstrate or indicate how they would respond. The key component of sample-based predictors is that they assess candidates on behaviors sampled from the criterion of interest. It is worthwhile to note that sample-based predictors fall along a continuum of how closely scenarios and response formats mirror actual situations on the job. This ranges from low-fidelity tools (e.g., situational judgment tests, biographical data) to high-fidelity tools (e.g., work sample, assessment centers) (Motowidlo, Dunnette, & Carter, 1990). When evaluating the cross-national generalizability of a selection instrument, the sign versus sample distinction is important because cultural factors may be more likely to influence the relevance of behaviors captured by a sample-based predictor or the

effectiveness of a behavioral response compared to a sign-based predictor (Briscoe, 1997; Lievens et al., 2003).

To address this gap, Lievens et al. (2015) examined the cross-national generalizability of a situational judgment test of integrity developed in the United States to a sample of employees in Spain. Results indicated that the integrity-based SJT was a significant predictor of self-reported counterproductive work behavior in both the U.S. and Spanish samples. Additionally, the magnitude of the relationships between the predictor and various criterion sub-dimensions (e.g., social conformity, driving violations, and perception of dishonesty norms) were not significantly different between the United States (.09-.26) and Spain (.15-.21). This primary study provides preliminary evidence for the cross-national generalizability of an SJT, which measures integrity, to Spain and other culturally similar countries.

Biographical data.

Similar to SJTs, biographical data is a standardized self-report selection method that asks applicants questions about prior behaviors and experiences (Mael, 1991). In contrast to SJTs, biographical data asks candidates about historical, rather than hypothetical behavior. At present, three published primary studies have reported conflicting results regarding the cross-national generalizability of biographical data inventories.

First, Laurent (1970) found that the validity of two different biographical data inventories, empirically keyed in the United States, remained significant predictors of managerial performance in Norway, Denmark, and the Netherlands. The two inventories asked questions related to “home and family background, education, vocational planning and experience, finances, leisure time activities, health history, and social and community relations” (Laurent, 1970, p. 418). In a similar study, Hinrichs, Haanperä, and Sonkin (1976) developed and

empirically keyed a biographical data inventory, which assessed work experiences, education, and temperament, in a sample of salespeople from Finland and Sweden and examined its generalizability to salespeople in Norway, France, Portugal, and the United States. The researchers found that the inventory generalized to culturally similar countries (i.e., Norway and the United States), but it was not a significant predictor of job performance in the more culturally distinct Latin Europe countries of Portugal and France (House et al., 2004). Third, Dalessio, Crosby, and McManus (1996) found that a biographical data inventory, which assessed work experiences, values, and interests, developed and empirically keyed in the U.S. for salespeople in the insurance industry remained a significant predictor of job performance when transported to the United Kingdom and Ireland.

Collectively, evidence suggests that biographical data inventories developed in one country (e.g., the United States) may only generalize to culturally similar countries and not to countries that are culturally dissimilar from the country in which it was developed. This is a tentative conclusion worth examining further, however, because all research has focused on European countries and in the study by Hinrichs and colleagues (1976), the sample size in the two countries where the validity did not generalize—France ($N = 44$) and Portugal ($N = 22$)—makes it difficult to disentangle the impact of statistical power from culture. Therefore, additional research is needed to better understand the cross-national generalizability of biographical data inventories.

In summary, research on the cross-national generalizability of selection instruments is limited. Most of it has focused on the generalizability of the two most commonly used selection instruments—cognitive ability and personality—to the European community. Recent work, however, has started to evaluate this question with other potentially useful selection instruments

(i.e., person-environment fit, SJTs) and in regions outside of Europe (i.e., East Asia). Given the expansion of the global economy to more culturally diverse countries, such as Brazil, Russia, India, and China, research is still in its infancy and more work is needed to evaluate the cross-national generalizability of selection instruments to a broader sample of countries. In the next section, I briefly review research on biographical data to explain why it may be useful for multinational organizations to consider including it in their global selection systems and provide evidence to support my hypothesis that carefully developed biographical data inventories will generalize to all country clusters examined.

A Review of Biographical Data

At this point, I briefly explained biographical data and the need for more research to evaluate its cross-national generalizability. In this section, I discuss biographical data more broadly. Specifically, I provide a definition, discuss its benefits, and detail how to develop and score inventories. Rather than providing a comprehensive literature review, this section is intended to provide readers a general overview of biographical data as it relates to the current research question. For more detailed information, see one of the many excellent literature reviews on the topic (e.g., Breaugh, 2009; Mumford et al., 2012; Schmitt & Golubovich, 2013).

What is biodata?

Despite slight variation in common definitions of biographical data (e.g., Hough, 2010; Nickels, 1994; Mael, 1991; Mount et al., 2000), there are three attributes common to all biographical data measures (Mumford et al., 2012). First, biographical data items ask about behaviors or experiences that previously occurred in an individual's life. Second, all test takers respond to the same set of questions (i.e., standardized). Last, responses are provided by the individual that is being assessed (i.e., self-report) (Mumford et al., 2012). Taken together,

biographical data is a standardized self-report assessment method that asks applicants questions about prior behaviors and experiences.

It is important to note that while this definition appears straightforward, in practice there are differences in how researchers operationalize “prior behavior and experiences”. For example, should biographical data measures include broader concepts such as temperament, skills, aptitudes, and attitudes (Mount et al., 2000)? Should measures focus on specific instances of behavior or more general behavioral patterns (Beatty, 2013)? Should items only focus on prior work behavior and experience or include other life domains (e.g., education, family, and volunteer experiences)? In order to better answer these questions, researchers have developed taxonomies of biographical data attributes to clarify what should and should not be considered biographical data.

Mael (1991) provides the most current and comprehensive taxonomy of biographical data item attributes. Based on an exhaustive review of previous work, he proposed a taxonomy with the following 11 attributes, with opposing anchors, and provided items to illustrate each positive dimension:

- *Historical* items focus on events that have already taken place or continue to take place (e.g., How old were you when you got your first job?). In contrast, *hypothetical* items ask about behavioral intentions or behavior in the future.
- *External* items emphasize prior behavior or experiences in actual situations (e.g., Have you ever been fired?) and *internal* items include attitudes, opinions, and reactions to events.

- *Objective* items focus on the recall of actual behaviors or experiences (How many hours did you study for your state licensing test?). *Subjective* items include one's interpretation of events (e.g., affective reactions, evaluation of one's abilities).
- *First-hand* items only ask about the direct knowledge of the examinee (e.g., How punctual are you about coming to work?). *Second-hand* items ask how other people (e.g., previous employer, teacher) would rate the examinee's performance or behavior.
- *Discrete* items focus on a single behavior or count of unique events (e.g., How many leadership positions did you hold during university?). *Summative* items ask about general or average tendencies over a period of time.
- *Verifiable* items, in contrast to non-verifiable items, can be corroborated by other sources (e.g., How long did you work for your previous employer?).
- *Controllable* versus *uncontrollable* items ask about events that are primarily determined by the examinee (e.g., How many times did it take you to pass the CPA exam?).
- *Equal access* emphasizes skills and events that are available to all examinees (Were you ever president of your high school class?) rather than *non-equal access* items, which may be influenced by external factors such as gender or socioeconomic status.
- *Job-relevant* emphasizes that items should appear to be job-relevant (e.g., How many hours did you bill to clients last year?) rather than ask about life events that are *not job relevant* or not clearly linked to job requirements.
- *Non-invasive*, compared to *invasive*, items do not ask about sensitive topics such as national origin, religion, political affiliation, or financial status (e.g., Have you ever served as a mentor to someone else?)

Mael (1991) argues that biographical data items should emphasize the first end (e.g., historical, external, objective) of the pole for each dimension. In addition, these attributes form three broader categories that help to define biographical data further. First, historical is a critical feature of biographical data. It is important for biographical data to focus on actual, rather than hypothetical, behavior. The second group emphasizes methodological attributes of biographical data items (i.e., external, objective, first-hand, discrete, and verifiable) and helps to ensure inventories obtain accurate behavioral information. The last group of attributes (i.e., controllable, equal access, job-relevant, and non-invasive) emphasizes legal considerations. Focusing on these poles for each dimension helps to minimize legal challenges from candidates. Collectively, these characteristics clarify the definition of biographical data and differentiate it from other self-report selection instruments such as personality, which typically focus on subjective and internal items.

Benefits of biographical data.

As previously mentioned, biographical data is a potentially useful tool that warrants additional research in the context of multinational personnel selection for a number of reasons. Specifically, when evaluating the usefulness of a selection instrument, organizations commonly consider the associated validity, cost, and legal implications (Ryan & Tippins, 2004). According to these criteria, biographical data has a number of favorable attributes.

First, research shows that biographical data inventories have generalized validity across companies when predicting a variety of organizational outcomes, including performance ratings, objective performance, tenure, creativity, and training success (Bliesener, 1996; Carlson, Scullen, Schmidt, Rothstein, & Erwin, 1999; Schmidt & Hunter, 1998). Across these outcomes, a biographical data inventory can be expected to have an uncorrected cross-validity of around .30 (Beatty, 2013). In addition, biographical data inventories have the potential to add incremental

validity over interviews (Dalessio & Silverhart, 1994), cognitive ability tests (Mael & Hirsch, 1993), and Big Five assessments (Mount, Witt, & Barrick, 2000; Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004) when predicting organizational outcomes. This suggests that, depending on the constructs assessed by an inventory, biographical data assessments have the potential to add value when included as an additional element in a selection system.

Second, while biographical data inventories tend to have moderately high development costs, administration costs tend to be low (Ryan & Tippins, 2004). Development costs tend to be higher because biographical data inventories typically require separate samples to develop and validate a scoring key. In addition, if creating an empirical key, it is common to write two to four times the number of items required for the final test form (Mumford et al., 2012). Despite these upfront costs, administration costs tend to be low because biographical data inventories are typically presented via paper and pencil or web-based techniques (Mumford et al., 2012). Therefore, in combination, biographical data inventories are reasonably cost-effective.

Last, from a legal standpoint, biographical data tends to have less adverse impact across racial and gender groups compared to other predictors such as GMA (Bobko & Roth, 2013; Bobko, Roth, & Potosky, 1999). Also, there have not been any legal challenges regarding their use for personnel selection (Terpstra, Mohamed, & Hethley, 1999) and candidates tend to have moderately favorable reactions to them (Hausknecht, et al., 2004). This last point is important because reactions to assessments have been linked to candidates' intentions to file legal complaints (Hausknecht et al., 2004). These findings suggest that there is minimal legal risk to using a validated biographical data inventory, even in a country with strict employment laws.

Collectively, research suggests that biographical data inventories have moderate validity when predicting a variety of outcomes, incremental validity over other commonly used selection

instruments, practical benefits, and minimal legal risk compared to other selection instruments. Based on these favorable properties and the lack of clarity in the available research literature, the cross-national generalizability of biographical data warrants additional research.

Theoretical foundation of biodata.

At the most basic level, biographical data inventories predict future job performance because of the principle of behavioral consistency. That is, the best predictor of future behavior is past behavior (Wernimont & Campbell, 1968). Specifically, scores relate to organizational outcomes because inventories assess candidates' prior job-relevant behaviors. This explanation, however, is deficient in that there is usually not an exact 1:1 mapping of behaviors assessed by a given biographical data inventory and the criterion space. In an effort to better explicate the mechanism underlying why biographical data works, Mumford, Stokes, and Owens (1990) developed the ecology model.

As noted by Mumford, Uhlman, and Kilcullen (1992), the overarching premise of the ecology model is that it "views the individual as an active, purposive entity who, through learning, cognition, and action, seeks to maximize personal adaptation in a world of shifting environmental opportunities" (p. 112). More specifically, this model proposes that environmental and hereditary resources influence individual differences early in one's life. Based on these differences (e.g., high GMA), a person begins the process of adapting to their environment.

This involves deciding what situations to enter based on the valence of the expected outcome, which is based on the needs and interests of the individual. In addition, individual differences influence what situations people enter and how well they perform. Based on how they perform in a situation, a person will develop new KSAOs, which will influence the valence of and performance in future situations. Eventually, this on-going process of entering into new

situations and developing additional skills results in a consistent pattern of behaviors. Ultimately, biographical data inventories capture information about these behavioral patterns and these patterns or correlates of these patterns influence candidates' future behavior on-the-job (Dean, Russell, & Muchinsky, 1999). That is, the ecology model helps to better understand why biographical data works and how past behavior predicts future behavior. Additionally, this theoretical framework provides guidance when developing biographical data items. Namely, biographical data inventories will be effective to the extent that they capture relevant behavioral patterns that are more recent in time (Mumford et al., 2012).

Development approaches.

While the ecology model provides general guidance on item development, it is helpful to review content development frameworks and strategies. At the broadest level, there are two item generation frameworks when developing a biographical data inventory for personnel selection (Mumford et al., 2012). The worker-oriented approach begins by identifying, via job analysis or competency modeling, the KSAOs believed to influence job performance. Next, the research team operationally defines how these KSAOs are likely to be manifested in candidates' prior life histories. Then, the researchers write items to tap these prior behaviors and experiences. For example, if Conscientiousness is important for a job, an item could ask about how many times a candidate failed to fulfill school or work responsibilities on time. In contrast, a work-oriented approach focuses on past behaviors that are directly relevant to the criterion domain. For example, if strategic planning is important for job performance, an item may ask, "In the past year, how many times were you involved in defining the key goals of a work group?".

When determining what framework to use to guide item development, both approaches have been used to create instruments with adequate reliability and criterion-related validity

(Mumford et al., 2012). Therefore, it is important to understand the intended purpose and audience for the assessment. Work-oriented approaches tend to result in items with high face validity but it assumes that candidates have prior experience in the work domain. If this assumption is not tenable, then a worker-oriented approach is likely to be more appropriate because items elicit information about candidates' standing on KSAOs believed to underlie successful job performance.

After deciding on an item generation framework, researchers must determine the technique(s) they will use to guide item development. Broadly speaking, there are three strategies, or sources of information, researchers use to write items (Mumford et al., 2012). First, work analysis information uses standard job analysis (e.g., functional job analysis, critical incidents method) and competency modeling techniques to identify job-relevant behaviors, and then write items to reflect these behaviors in similar situations that candidates may have experienced in their past. A second approach to item development relies on substantive understanding and relevant theories. When using this approach, researchers examine prior research and theory on the construct of interest to understand how it influences peoples' behaviors and experiences. Based on this information, researchers can develop items that are relevant to the intended audience. Last, researchers can use qualitative methods to generate biographical data items. Specifically, researchers can study the life histories of those who have performed exceptionally well or poor in the job of interest. This can involve interviewing job incumbents or asking them to write life history essays. The result of this qualitative research is a rich source of information to draw on when developing items. In terms of best practices, researchers have successfully used all three approaches when developing biographical data inventories (Mumford et al., 2012). Therefore, a research team may considered one or more

approaches as part of their item generation strategy while weighing practical considerations (e.g., a job analysis was already conducted) to determine the best approach².

After developing items, it is important for researchers to screen them. Specifically, Mumford and colleagues (1996) note that it is important to evaluate items for: (1) relevance to the construct of interest, (2) contamination with other constructs, (3) social desirability, (4) bias regarding equal access, (5) potential for faking, (6) invasion of privacy, and (7) controllability of the behavior. Additionally, in the context of cross-national selection, it is important to review items to ensure that they are relevant across the countries in question and do not use colloquial language, idioms, or terms unique to one culture (Schaffer & Riordan, 2003). Researchers should modify or exclude problematic items. Once the items are finalized, researchers can administer them in a sample of respondents who are likely to be similar to the test-takers. The purpose of this is to develop a scoring key and validate the instrument before deploying it with candidates. Similar to writing biographical data items, scoring the resulting responses involves many decisions, which are discussed in the next subsection.

Scoring approaches.

After developing and administering a pool of biographical data items, the next step for researchers is to create a summary score of a candidate's responses. Since biographical data items assess behavior rather than ability, they do not usually have a single "right" answer. Therefore, researchers must make decisions about what items to include in an inventory and how to weigh each response. In general, there are four broad methods to generate summary scores: (1) empirical keying, (2) rational scaling, (3) factorial scaling, and (4) subgrouping (Mumford &

² In the present study, both biographical data inventories were developed with a worker-oriented approach and relied on data from work analyses and theory to generate items.

Owens, 1987). In this section, I briefly review and compare these approaches in the context of personnel selection.

First, empirical keying approaches select and weight items based on their relationship with a criterion measure. Items that exhibit high correlations with the criterion are included in the final inventory. As noted by Mumford and Owens (1987), a correlation of .10 to .15 has been found to be a useful cut point when determining what items to include in the final scoring key. Including items with smaller relationships with the criterion has the potential to disproportionately increase error variance, which will impact the reliability and criterion-related validity of the instrument.

Test developers have commonly used empirical approaches because they maximize the proportion of variance accounted for in an outcome variable (Mumford et al., 2012). This approach, however, is not without limitations. First, the utility of an empirical scoring key is heavily dependent on the adequacy and quality of the criterion measure. The operational validity of an empirically scaled biographical data inventory will be attenuated when criterion measures are deficient and/or contaminated. Additionally, since keys are developed in a specific sample, they have the potential not to generalize to other situations or criterion measures, especially when the criterion measure in the new situation differs from the original sample (Mumford & Owens, 1987). Lastly, numerous researchers have noted that empirical keys are atheoretical (e.g., Bergman, Drasgow, Donovan, Henning, & Juraska, 2008; Schmitt & Golubovich, 2013). This is criticism is unjustified because researchers use an approach consistent with scientific theory when they rely on the aforementioned sources of information (e.g., job analysis) when writing items and make hypotheses about which response options are likely to be predictive of performance (Cucina & Walmsley, 2015). Assuming researchers address these concerns,

empirical keys have shown to produce impressive validities ($\sim .30$) and retest reliabilities ($\sim .80$) (Mumford et al., 2012). It is important to note that developing an empirical key requires access to criterion data, a sample of at least several hundred people, and cross-validation to ensure the stability of the key (Cucina, Caputo, Thibodeaux, & Maclane, 2012).

Second, rational scaling approaches select items and specify weights based on experts' judgment and their knowledge of the construct/content domain. When using this approach, researchers typically write items to tap specific constructs believed to influence the criterion of interest and weigh response options based on the level of the construct they reflect. Items that tap the same construct are summed together to create dimension sub-scores, which can then either be analyzed separately or combined to predict the criterion of interest. The benefit of a rational approach is that scores are not tied to a specific criterion or sample. This implies that rationally developed scales are more likely to generalize to new situations because they assess constructs. A rationally developed biographical data inventory's validity is determined by the relevance of the constructs to the criterion, rather than being built into the scale as is the case with empirical approaches (Mumford & Owens, 1987; Mumford et al., 2012). Therefore, a rationally developed biographical data inventory will only be effective to the extent that the constructs assessed influence criterion behaviors (Binning & Barrett, 1989). Additionally, since rational keys rely on expert judgment, it is easier for candidates to discern the "correct" response for a given item and artificially inflate their scores when compared to an empirical key (Kluger, Reilly, & Russell, 1991).

Third, factorial scaling approaches use factor analytic methods (e.g., exploratory factor analysis) to identify meaningful psychological dimensions underlying responses to items on a biographical data inventory. The resulting solutions are then evaluated against common criteria

such as simple structure and theoretical clarity. Items that have low loadings or load on multiple factors are excluded from the final scale. The remaining items are weighted, usually based on factor loadings or unit weights, and summed to create dimension scores. Research shows that factorial scaling approaches typically exhibit high internal consistency reliability ($\sim .80$) but similar to rational scaling, factorial scaling does not ensure that the resulting inventory will predict performance (Mumford & Owens, 1987). Factorial scaling has many of the same advantages as rational scaling because they both focus on the constructs underlying a scale. Rational scaling, however, requires a prior knowledge of the constructs assessed by the inventory (Mumford et al., 2012). Therefore, in general, researchers have more commonly used rational rather than factorial scaling.

Last, subgrouping focuses on identifying clusters of people based on a similar pattern of responses. This approach involves: (1) using clustering algorithms (e.g., latent class analysis) to identify groups of people who tend to respond similarly, (2) determining the number of meaningful clusters to retain, and (3) describing the unique characteristics in each cluster (Beatty, 2013). Personnel selection researchers infrequently use this approach because scores identify a candidate's similarity to other people rather than predict his or her future performance. Additionally, clustering algorithms do not always classify all individuals and may predict that some people belong to multiple groups. Therefore, this approach is problematic for use in operational high-stakes selection.

In the context of personnel selection, researchers have commonly used an empirical, a rational, or a hybrid (i.e., empirical + rational) scoring approach rather than factorial or subgrouping procedures (Cucina et al., 2012). While some studies have compared these approaches (e.g., Mitchell & Klimoski, 1982), Cucina and colleagues (2012) published the most

extensive study on the effectiveness of a variety of different keying methods within each broad scoring category (empirical, rational, hybrid) on the criterion-related validity of biographical data inventories. They found that, in general, empirical scaling methods had the highest validities when development sample sizes were below about 1,600³ and rational weights performed worst.

In terms of specific empirical methods (e.g., vertical percent method, biserial correlation, point biserial), the approaches were very highly intercorrelated ($r \geq .90$) and yielded similar cross-validities. Based on this finding, Cucina and colleagues (2012) recommend the point biserial raw weights method. This approach, which is used in the current study, dichotomously scores individual response options within each item to indicate if a participant endorsed a specific response. Then, response option weights are simply the product of the point biserial correlation with the criterion and endorsement (0, 1), assuming the outcome is continuous. The equation below illustrates the process of creating a total score with this approach (Dean, Russell, & Farmer, 2002).

$$\text{Candidates biographical data score} = (r_{1,1} * ro_{1,1}) + (r_{1,2} * ro_{1,0}) + \dots (r_{n,n} * ro_{n,n})$$

Where:

$R_{N,N}$ = correlation between item 1's response option 1 and the criterion in the development sample, and

$RO_{N,N}$ = equals 0 if a candidate did not chose item 1's response option B, or 1 if the respondent chose item A, response option B.

³ When samples were larger than about 1,600, a hybrid keying approach with stepwise regression weighting resulted in the highest cross-validities. Given that the development samples in the current study range between 500 and 1,000, a hybrid keying approach was not utilized.

In this section, I provided an overview of biographical data research. Specifically, I defined biographical data and discussed attributes that differentiate it from other self-report measures, such as personality. Next, I discussed the practical benefits of biographical data including validity, cost, and legal implications. Then, I reviewed approaches to develop and empirically key biographical data inventories. Based on current best practices, the present study used a worker-oriented approach, relied on data from work analyses and theory to generate items, and scored response options with the point biserial raw weights method to evaluate the cross-national generalizability of two biographical data inventories.

Conclusion

In summary, I have discussed the strategic importance of personnel selection for organizations, the unique challenges of personnel selection for multinational organizations, the cross-national generalizability of selection instruments, and relevant research on biographical data. Based on this review, the present studies will investigate the cross-national generalizability of two biographical data inventories developed in the United States to a diverse sample of country clusters in Study 1 (Anglo-Saxon, Confucian Asia, Germanic, Latin Europe, and Southern Asia) and Study 2 (Confucian Asia, Eastern Europe, Germanic, Latin America, and Southern Asia). These two studies use the archival data of employees from two occupational groups—managers (Study 1) and manufacturing technicians (Study 2)—from one multinational consumer goods company. Based on the available evidence related to the generalizability of selection instruments and biographical data content development best practices, I hypothesize that empirically keyed worker-oriented biographical data inventories, which use rigorous job analysis methods and theory as the foundation for content generation, will generalize to all country clusters examined. If supported, the findings would extend current research and suggest

that biographical data instruments have the potential to enhance the value of multinational organizations' global selection systems.

Method - Study 1

Participants

Participants were 2,763 job incumbents that worked for a single large multinational consumer goods company. All participants had lower or middle level managerial positions and worked in a variety of organizational functions (e.g., Sales, Design, and Human Resources). The sample had slightly more males (55.2%) compared to females (43.5%) and over 90% of participants had at least a four year university degree. The majority of the sample was from the United States ($N = 1,616$) but participants also came from Belgium ($N = 71$), Canada ($N = 106$), China ($N = 113$), France ($N = 59$), Germany ($N = 83$), India ($N = 79$), Italy ($N = 106$), Japan ($N = 308$), and the United Kingdom ($N = 222$). Participants self-selected the country in which they had spent the majority of their lives. Including the United States, 10 countries were represented in the sample. To increase statistical power, country clusters were created based on the GLOBE framework (House et al., 2004). Excluding the United States, this resulted in five clusters: Anglo-Saxon (Canada, United Kingdom), Confucian (China, Japan), German (Belgium, Germany), Latin Europe (France, Italy) and Southern Asia (India).

As shown in Table 1, there were a few minor differences in the distribution of gender and educational attainment but function was similar across country clusters. That is, the organization had each functional area (e.g., Design, Marketing, Human Resources) in each region, rather than consolidating some strategically important functions (e.g., Marketing, Research & Development) in their global headquarters and outsourcing back-office functions (e.g., Information Technology,

Human Resources) to specific, low-cost regions. Based on this structure, the job complexity and KSAOs required to succeed as a manager were comparable across country clusters.

Measures

Job performance.

Each employee's immediate supervisor assessed his or her job performance with 41 behaviorally-based Likert-type items, which were organized into nine performance competencies as identified through a separate managerial global competency modeling project. As previously noted, the purpose of competency modeling is to identify the KSAOs needed to execute an organization's strategy. In contrast to job analysis, competency modeling seeks to identify behavioral dimensions that are common across multiple jobs in an organization. That is, the objective of competency modeling is to create one centralized KSAO profile across a large group of jobs. An intermediate step in this process is ensuring there is a common job across countries—a key requirement when a multinational organization elects to pursue a global or transnational selection strategy (i.e., a standardized selection system across geographies).

The competency modeling project used interviews, focus groups, and questionnaires to identify the competencies critical for job performance within the organization and confirmed that the fundamental KSAOs required to perform managerial jobs were similar across geographies. This is not to say that all jobs required the exact same skills. Rather, there was a foundational set of KSAOs for all managers (i.e., the competency model) aligned with the overall organizational strategy but job-specific or technical knowledge varied across functions (e.g., Legal vs. Design). The resulting nine dimensions for the job performance measure included KSAOs such as leadership, innovation, collaboration, and decision-making. Four to five items were used to assess each competency and each item was rated on a 5-point scale ranging from “Weak” to

“Exceptional”. Additionally, there was an escape option for supervisors to indicate if they had not observed a given behavior.

A confirmatory factor analysis, which specified a model with the average score (e.g., average of five items related to leadership competency responsibilities -1. envisions, 2. engages, 3. energizes, 4. enables, and 5. executes) on each of the nine competencies loading on a general performance factor, indicated adequate fit ($\chi^2(27) = 699.21$, RMSEA = 0.09, TLI = 0.95, SRMR = 0.03; Hu & Bentler, 1999). Coefficient alpha for the entire performance rating form indicated excellent internal consistency ($\alpha = .96$). In combination, these sources of statistical evidence suggested that there was a general factor of job performance (Viswesvaran, Schmidt, & Ones, 2005). Therefore, a composite performance criterion measure was created by averaging items within a competency and summing across each of the nine competencies to create a unit-weighted (at the competency level) overall job performance score.

Biographical data inventory.

A 206-item biographical data inventory was administered to job incumbents as part of a concurrent validation study. The inventory items were developed to assess behaviors and experiences associated with each of the nine aforementioned dimensions of managerial performance (e.g., leadership, innovation, and collaboration). Information gathered from the global managerial competency modeling project plus relevant psychological research served as input when writing biographical data items. An example item written to assess the leadership competency was, “How many times have you been involved in defining the key goals of a group (e.g., service group, department, organization, etc.)?”.

Items were reviewed by a panel of industrial-organizational (I-O) psychologists and evaluated to ensure the content was globally appropriate (Mumford et al., 1996; Schaffer &

Riordan, 2003). This item review was critical because, as previously mentioned, it was not possible to test for measurement invariance for the biographical data inventory because it was multidimensional, it used a variety of different response formats (e.g., Likert-type, dichotomous), and many countries did not have a sufficient sample size. The first two characteristics are common to many biographical data inventories and make it difficult to test for measurement invariance because the data is likely to violate the assumption of multivariate normality. In this situation, Lievens and colleagues (2015) note that it is more important to focus on conceptual equivalence, or the extent to which items and constructs have similar meanings across countries. Therefore, the team of I-O psychologists evaluated items for: (1) relevance to the construct of interest, (2) contamination with other constructs, (3) social desirability, (4) bias regarding equal access, (5) potential for faking, (6) invasion of privacy, and (7) controllability of the behavior. Additionally, in the context of cross-national selection, items were reviewed to ensure that they were relevant across all country clusters and did not use colloquial language, idioms, or terms unique to one culture (Schaffer & Riordan, 2003).

An empirical scoring key for the biographical data inventory was developed and cross-validated in the sample of U.S. participants. Two-thirds of the sample ($N = 1,074$) was randomly selected as the development group and one-third ($N = 542$) was used to cross-validate the scoring key. This approach is consistent with similar efforts for developing a biographical data inventory (Cucina et al., 2012). I empirically keyed and scored the biographical data inventory with a point biserial raw weight scoring approach and retained items with response options that had a statistically significant point biserial correlation with the composite criterion $\geq |.10|$ (Cucina et al., 2012). As noted by Mumford and Owens (1987), a correlation of .10 to .15 has been found to be a useful cut point when determining what items to include in the final scoring key. Including

items with smaller relationships with the criterion has the potential to disproportionately increase error variance, which will impact the reliability and criterion-related validity of the instrument.

This resulted in a final biographical data inventory with 45 items.

Procedure

A subset of employees from the multinational organization were invited to participate in the research study as part of a concurrent global validation study. Participants were selected based on a stratified random sampling plan to ensure representativeness of the organization's management population by key demographic variables (e.g., gender, function, geography, and race [U.S. only]). Selected employees were invited to participate in the research study via email. The invitation emphasized that participation was voluntary and confidential and demographic information for each participant would be obtained from the organization's human resources management system (HRMS).

Willing participants were directed to a website that hosted the research form in their native language or English. A professional translation agency completed all translations, which native speakers within the organization reviewed and compared to the English version. Any discrepancies were noted by the reviewers and addressed via consensus discussion. In order to minimize translation costs during the validation study, the organization decided to translate the research form into a country's native language for the validation study based on historical hiring data so that countries who hired more employees received translated research forms. Then, additional translations would be carried out for other languages based on business need, after finalizing the biographical data inventory. It is important to note that researchers commonly develop two to four times the number of biographical items than desired for the final assessment. Therefore, rather than translating all 206 biographical data research items into each language, the

organization elected to wait to translate the assessment into some languages until after determining the final set of items. Furthermore, this practical decision, while not ideal, does not affect the scientific rigor of the research.

In study 1, participants from Belgium, Canada, India, Italy, the United Kingdom, and the U.S. completed the inventory in English. Participants from China (Simplified Chinese), France, Germany, and Japan completed the form in their native language. It is worth acknowledging that some participants completed the assessment in English, as opposed to their native language. English language proficiency, however, is required for managerial jobs within the organization and is evaluated before employees are hired. Therefore, it was not problematic to administer the biographical data form to participants in their non-native language.

Concurrent with this data collection effort, research-only job performance ratings were gathered from each participant's immediate supervisor via an online survey. Prior to completing the job performance measure, supervisors received rater error and performance dimension training to reduce common rater biases and improve accuracy (Woehr & Huffcutt, 1994). Additionally, the instructions for the rating form emphasized the confidential nature and research purposes of the ratings and instructed supervisors to rate honestly. Employees and supervisors completed all research measures during normal business hours and were compensated for their time with their standard salary.

Table 1*Study 1 Demographic Information by Country Cluster*

Demographic Variable	<u>Overall</u>		<u>U.S (Cross)</u>		<u>U.S. (Dev.)</u>		<u>Anglo-Saxon</u>	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
<u>Gender</u>								
Male	1526	55.2	285	52.6	582	54.2	175	53.4
Female	1203	43.5	249	45.9	479	44.6	151	46.0
Prefer not to answer	34	1.2	8	1.5	13	1.2	2	0.6
<u>Education</u>								
Not a high school grad	6	0.2	0	0.0	0	0.0	2	0.6
GED	5	0.2	0	0.0	0	0.0	1	0.3
High school or equivalent	99	3.6	22	4.1	28	2.6	19	5.8
Associate's or equivalent	103	3.7	18	3.3	40	3.7	9	2.7
University degree (Bachelor's)	1408	51.0	299	55.2	602	56.1	178	54.3
Master's degree or equivalent	859	31.1	142	26.2	294	27.4	74	22.6
Advanced degree (e.g., PhD, MD, JD)	257	9.3	60	11.1	106	9.9	42	12.8
Other	26	0.9	1	0.2	4	0.4	3	0.9
<u>Function</u>								
Sales	476	17.2	95	17.5	190	17.7	65	19.8
Consumer & Market Research	110	4.0	25	4.6	39	3.6	16	4.9
Design	52	1.9	15	2.8	28	2.6	3	0.9
Employee Relations	61	2.2	6	1.1	24	2.2	6	1.8
Finance & Accounting	227	8.2	33	6.1	100	9.3	25	7.6
Human Resources	63	2.3	12	2.2	26	2.4	4	1.2
Information Technology	148	5.4	28	5.2	60	5.6	26	7.9
Legal	16	0.6	3	0.6	3	0.3	3	0.9
Marketing	154	5.6	20	3.7	42	3.9	23	7.0
Engineering	693	25.1	142	26.2	250	23.3	76	23.2
Research & Development	763	27.6	163	30.1	312	29.1	81	24.7

Table 1 (continued)*Study 1 Demographic Information by Country Cluster*

Demographic Variable	<u>Overall</u>		<u>U.S (Cross)</u>		<u>U.S. (Dev.)</u>		<u>Anglo-Saxon</u>	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Belgium								
Canada							106	32.3
China								
France								
Germany								
India								
Italy								
Japan								
United Kingdom							222	67.7
United States			542	100.0	1,074	100.0		
<u>Race (U.S. only)</u>								
Asian	99	5.6	24	4.4	48	4.5		
Black	104	5.9	29	5.4	74	6.9		
Hispanic	51	2.9	17	3.1	33	3.1		
White	1205	67.9	351	64.8	724	67.4		
Other	5	0.3	1	0.2	4	0.4		
Missing	311	17.5	120	22.1	191	17.8		

Table 1 (continued)*Study 1 Demographic Information by Country Cluster*

Demographic Variable	<u>Confucian</u>		<u>Germanic</u>		<u>Latin Europe</u>		<u>Southern Asia</u>	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
<u>Gender</u>								
Male	228	54.2	101	65.6	92	55.8	63	79.7
Female	186	44.2	52	33.8	70	42.4	16	20.3
Prefer not to answer	7	1.7	1	0.6	3	1.8	0	0.0
<u>Education</u>								
Not a high school grad	1	0.2	1	0.6	2	1.2	0	0.0
GED	0	0.0	3	1.9	1	0.6	0	0.0
High school or equivalent	6	1.4	9	5.8	13	7.9	2	2.5
Associate's or equivalent	31	7.4	2	1.3	1	0.6	2	2.5
University degree (Bachelor's)	253	60.1	10	6.5	52	31.5	14	17.7
Master's degree or equivalent	112	26.6	101	65.6	84	50.9	21	26.6
Advanced degree (e.g., PhD, MD, JD)	6	1.4	25	16.2	9	5.5	31	39.2
Other	12	2.9	3	1.9	3	1.8	9	11.4
<u>Function</u>								
Sales	66	15.7	23	14.9	31	18.8	6	7.6
Consumer & Market Research	16	3.8	3	1.9	3	1.8	8	10.1
Design	2	0.5	2	1.3	2	1.2	0	0.0
Employee Relations	15	3.6	0	0.0	10	6.1	0	0.0
Finance & Accounting	21	5.0	24	15.6	15	9.1	9	11.4
Human Resources	11	2.6	3	1.9	7	4.2	0	0.0
Information Technology	14	3.3	7	4.5	8	4.8	5	6.3
Legal	2	0.5	1	0.6	2	1.2	2	2.5
Marketing	23	5.5	21	13.6	21	12.7	4	5.1
Engineering	111	26.4	40	26.0	50	30.3	24	30.4
Research & Development	140	33.3	30	19.5	16	9.7	21	26.6

Results - Study 1

Descriptive Statistics

Table 2 provides descriptive statistics for managers on the biographical data inventory and job performance measure by country cluster. Preliminary analyses indicated the mean score on the biographical data inventory is significantly different between country clusters, *Welch's* $F(6, 564.15) = 44.92, p < .001, \omega^2 = .08^4$. Games-Howell⁵ post hoc tests indicated the mean score in the United States cross-validation sample ($M = 1.14$) is significantly greater than the mean score in the Germanic ($M = 0.63; d = 0.34$) or Confucian Asia ($M = -0.14; d = 0.80$) samples. Also, the mean score on the job performance measure is significantly different between the country clusters, *Welch's* $F(6, 566.78) = 3.97, p < .001, \omega^2 = .01$. Games-Howell post hoc tests indicated the mean score in the United States cross-validation sample ($M = 30.48$) is significantly greater than the mean score in the Confucian Asia sample ($M = 29.51; d = 0.19$). Given the large sample size for many of the groups, it is important to emphasize that the overall effect size for the impact of country cluster on the biographical data inventory and job performance measure are small and trivial, respectively. This suggests that these statistically significant differences were of minor practical importance. Furthermore, they are consistent with prior research, which has found that individuals from East Asian countries tend to use the midpoint of Likert scales more frequently compared to individuals from the U.S. (Chen et al., 1995) and biographical data

⁴ Since Levene's Test for Homogeneity of Variances was significant for both measures, Welch's F is reported, which adjusts the F statistic and the residual degrees of freedom for violations of the homogeneity of variance assumption.

⁵ Games-Howell post hoc tests are appropriate if there are unequal variances between groups and/or unequal group sizes (Field, 2005).

inventories commonly exhibit moderate adverse impact across racial and gender groups (Bobko & Roth, 2013).

The skewness and kurtosis values within each country cluster (Table 2) provide evidence that scores on the biographical data inventory and job performance measure are approximately normally distributed – a key assumption for parametric statistical tests. All skewness values are less than 1.0 and all kurtosis values, with one exception, are less than 1.0. Since the sample size in many country clusters was large (>200), the associated standard errors tended to be small. In this situation, the formal statistical test of dividing the value by the standard error (distributed as z) tends to be sensitive to even small departures from normality. Field (2005) notes that with large samples, it is more important to visually examine the shape of the distribution rather than calculate the test statistic and interpret the result. Figure 1 provides the visual plot of scores for the biographical data inventory and job performance measure by country cluster. This provides further support for the relative normality of each variable by country cluster.

The job performance measure (Table 2) had excellent internal consistency reliability within each country cluster ($\alpha > .90$). Coefficient alpha was not calculated for the biographical data inventory because the empirical scoring key was multidimensional and measures of internal consistency are inappropriate for this assessment method (Mumford et al., 2012). Rather, retest reliability is a more appropriate measure of reliability. Unfortunately, due to practical constraints (i.e., time required, expense), it was not possible to re-administer the same measure to all or a subset of job incumbents during the validation study to calculate retest reliability. Evidence from other studies, however, suggests that empirically keyed biographical data inventories produce impressive retest reliabilities after periods ranging from two months ($r_{RT} = .91$; Erwin & Herring, 1977) to 19 months ($r_{RT} = .85$; Chaney & Owens, 1964) (as cited in Mumford et al., 2012).

Table 2*Study 1 Descriptive Statistics by Country Cluster*

Country Cluster	Measure	<i>N</i>	<i>Mean (SD)</i>	<i>Skew (SE)</i>	<i>Kurtosis (SE)</i>	α
Anglo-Saxon	Biodata	328	1.27 (1.49)	-0.35 (.13)	-0.39 (.27)	
	Job Performance	328	30.92 (4.60)	-0.28 (.14)	0.51 (.27)	0.96
Confucian Asia	Biodata	421	-0.14 (1.59)	-0.10 (.12)	-0.83 (.24)	
	Job Performance	421	29.51 (4.86)	-0.24(.12)	0.64 (.24)	0.96
Germanic	Biodata	154	0.63 (1.36)	0.10 (.20)	-0.60 (.39)	
	Job Performance	154	30.58 (3.91)	-0.13 (.20)	1.26 (.39)	0.94
Latin Europe	Biodata	165	1.34 (1.44)	-0.35 (.19)	-0.13 (.38)	
	Job Performance	165	31.13 (4.43)	-0.44 (.19)	0.54 (.38)	0.96
Southern Asia	Biodata	79	0.82 (1.29)	-0.21 (.27)	-0.64 (.54)	
	Job Performance	79	30.95 (4.69)	0.22 (.27)	0.80 (.53)	0.96
United States (CV)	Biodata	542	1.14 (1.60)	-0.24 (.11)	-0.25 (.21)	
	Job Performance	542	30.48 (5.16)	-0.60 (.11)	0.08 (.21)	0.96
United States (Dev)	Biodata	1,074	1.21 ((1.61)	-0.36 (.08)	-0.26 (.15)	
	Job Performance	1,074	30.49 (5.17)	-0.41 (.08)	0.16 (.15)	0.97

Note. *N* = sample size; *SD* = standard deviation; *SE* = standard error; α = coefficient alpha; CV = cross-validation sample; Dev = development sample.

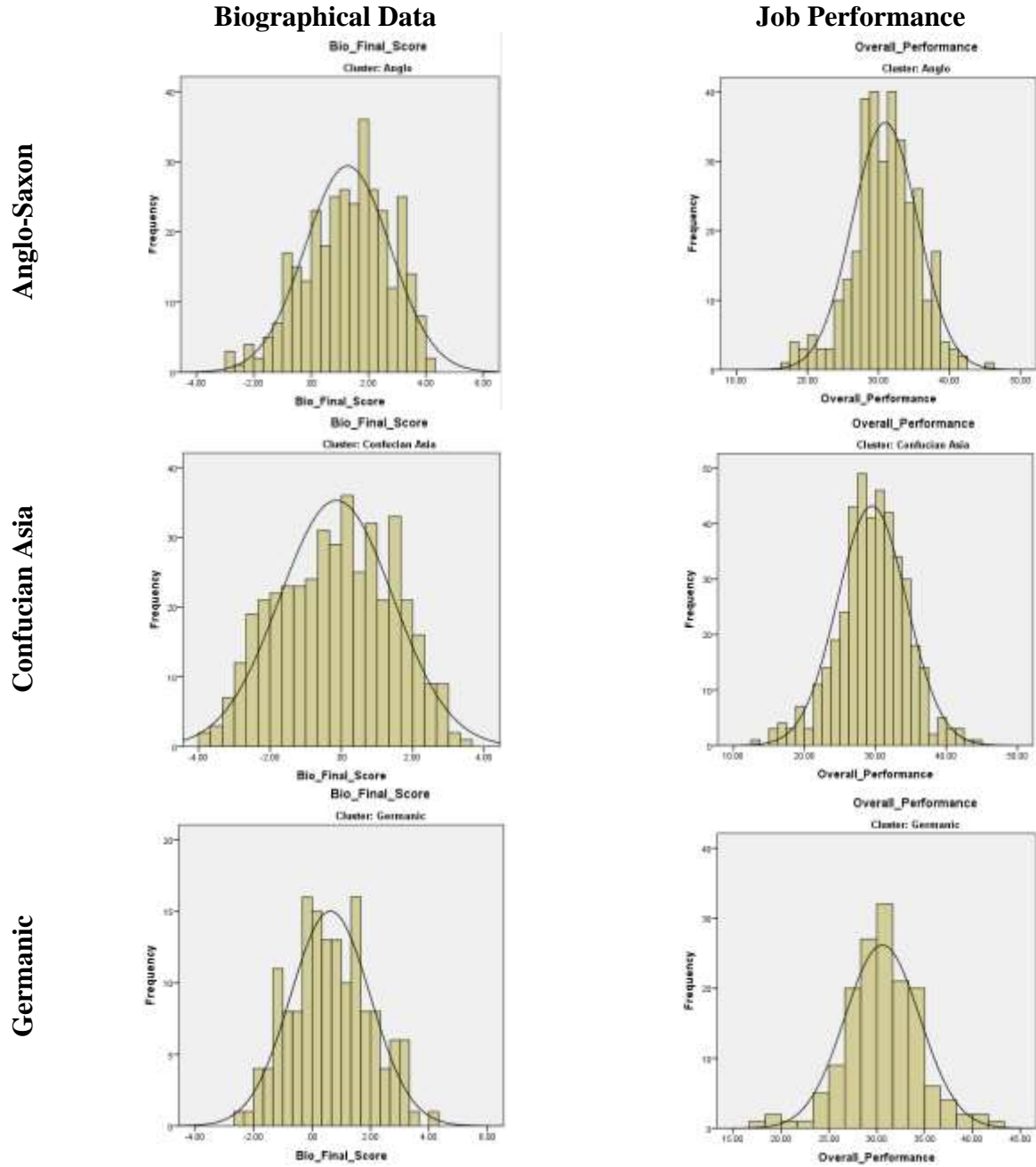


Figure 1. Study 1 Distribution of Variables by Country Cluster.

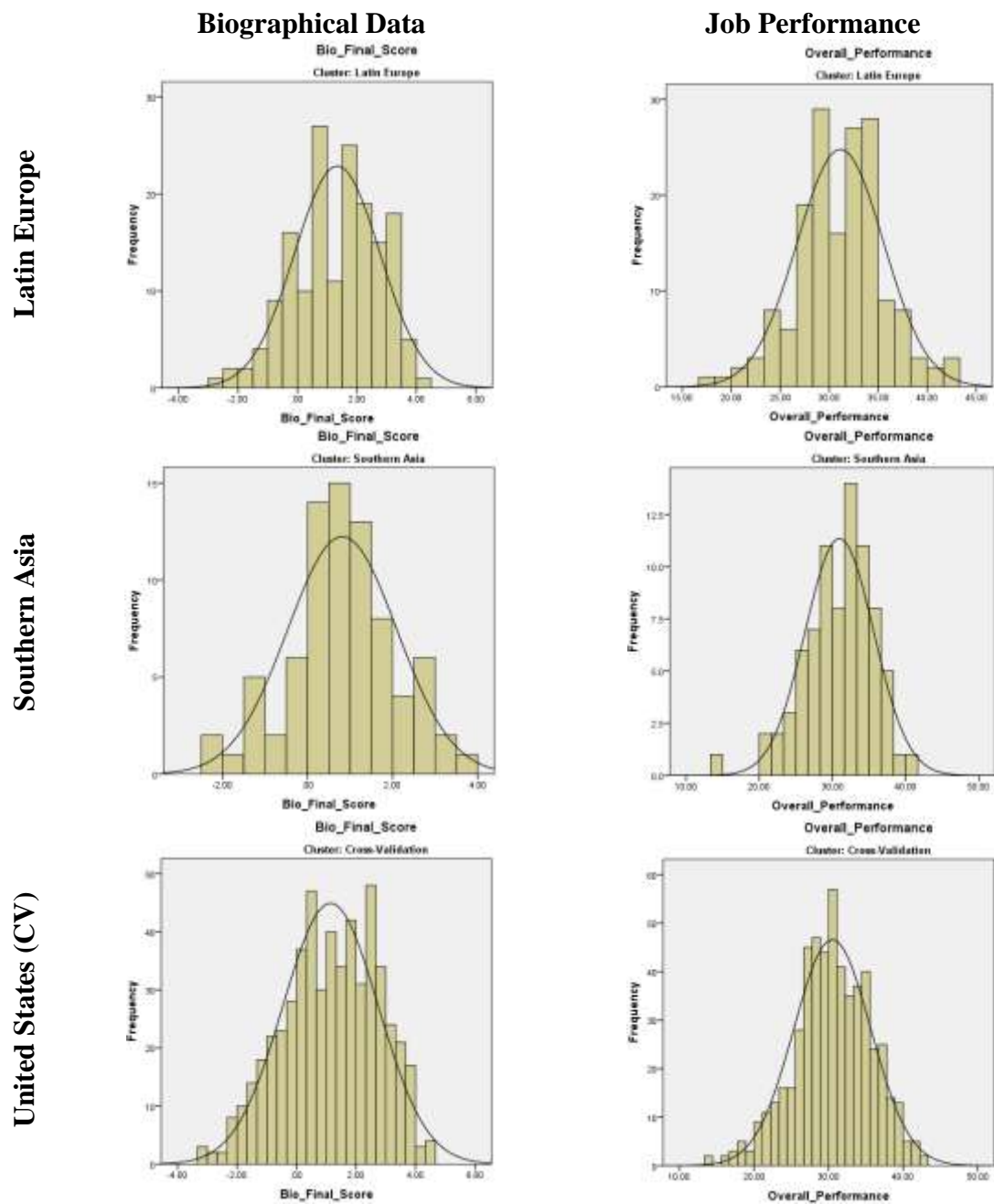


Figure 1. Study 1 Distribution of Variables by Country Cluster (continued).

United States (Dev)

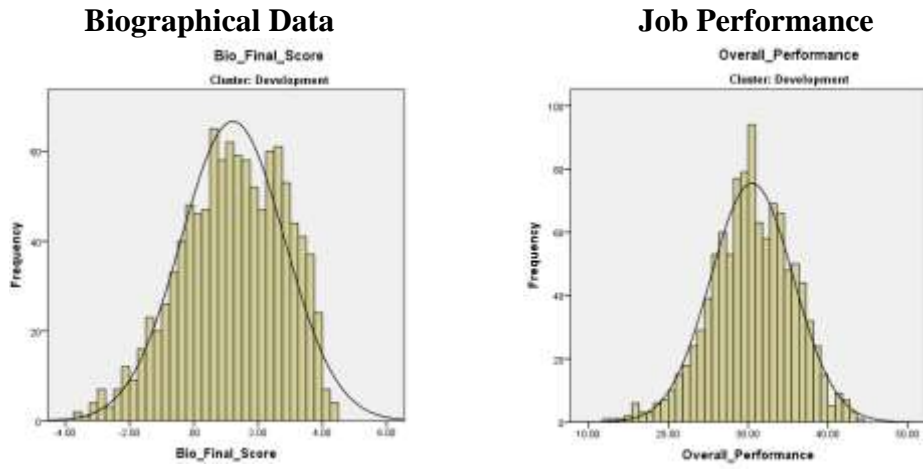


Figure 1. Study 1 Distribution of Variables by Country Cluster (continued).

Hypothesis Testing

To evaluate the cross-national generalizability of the biographical data inventory, the U.S. managerial sample was randomly divided into development (2/3) and cross-validation (1/3) samples. After creating summary scores for the predictor and criterion measures according to the approaches outlined in the Method section⁶, results indicated that the biographical data inventory was significantly related to job performance in the U.S. development sample, $r(1,072) = 0.39, p < .01$.

To ensure that the empirical scoring key and its relationship with the criterion measure did not capitalize on chance or the unique characteristics of the development sample, the same scoring key was applied to a separate U.S. holdout sample. When the biographical data inventory's empirical scoring key was applied to this U.S. cross-validation sample, it remained a significant predictor of job performance, $r(540) = 0.34, p < .01$. The shrinkage ($r = 0.05$) observed in the validity between the development ($r = 0.39$) and cross-validation ($r = 0.34$) sample was expected because the scoring key was based on data from a specific (development) sample of participants. Further, the magnitude of the relationship and degree of shrinkage observed in the U.S. cross-validation managerial sample is consistent with cross-validated empirically keyed biographical data inventories reported in previous research (Beatty, 2013; Cucina et al., 2012; Schmidt & Hunter, 1998).

Next, the same empirical scoring key was separately applied to respondents from each of the five country clusters to examine the validity of the U.S. empirical scoring key in each region. Given that some country clusters had relatively small sample sizes (e.g., Southern Asia, Germanic), it is important to note that the minimum sample size is 49 to have power equal .80

⁶ The biographical data inventory was empirically keyed and scored in the U.S. development sample with a point biserial raw weight scoring approach. Items included in the final key had a response option that had a statistically significant point biserial correlation with the composite criterion $\geq |.10|$.

when $\rho = .35$ with a 1-tailed test (Faul, Erdfelder, Buchner, & Lang, 2009). The country cluster with the lowest sample size was Southern Asia ($N = 79$). This suggests that all country clusters had a sufficient sample size to achieve adequate statistical power.

Results indicated that the inventory was significantly related to job performance in each of the clusters: (a) Anglo-Saxon, $r(326) = 0.35, p < .01$, (b) Confucian Asia, $r(419) = 0.31, p < .01$, (c) Germanic, $r(152) = 0.29, p < .01$, (d) Latin Europe, $r(163) = 0.26, p < .01$, and (e) Southern Asia, $r(79) = 0.33, p < .01$. In addition, the magnitude of the validity coefficient observed in each country cluster appeared to be similar to the validity observed in the U.S. cross-validation sample ($r = 0.34$). Collectively, these results provide preliminary support for the cross-national generalizability of the biographical data inventory because summary scores on the biographical data inventory, which were derived from the same U.S.-based empirical scoring key, were significantly related to job performance in the U.S. cross-validation sample and each country cluster. Refer to Table 3 and Figure 2 for detailed information about the criterion-related validity of the biographical data inventory by country cluster in the managerial sample.

Then, an omnibus test of the equality of independent correlations was used to formally examine the hypothesis that the magnitude of the observed validity in the U.S. cross-validation sample was not significantly different from the magnitude in each of the other country clusters (Chen & Popovich, 2002). This test is important to show that not only does the biographical data inventory significantly relate to job performance in each country cluster, but that the strength of the relationship is similar in each region. That is, the biographical data inventory is similarly effective when deployed in each geographical region.

This χ^2 test of homogeneity evaluates the null hypothesis that the correlation in each independent sample is equal (i.e., $\rho_1 = \rho_2 = \dots = \rho_k = \rho$). The alternative hypothesis says that at

least one correlation is different from ρ . Using the formula provided by Chen and Popovich (2002), this test indicated that there were no significant differences between the validity coefficients in the managerial sample by country cluster, $\chi^2(5) = 1.56, ns$. That is, given the observed data, it is extremely unlikely that the null hypothesis is true. This omnibus test provides further evidence to support the cross-national generalizability of the biographical data inventory, developed in the U.S., to a diverse sample of Anglo-Saxon (Canada, United Kingdom), Confucian Asia (China, Japan), Germanic (Belgium, Germany), Latin Europe (France, Italy), and Southern Asia (India) country clusters.

After finding that the biographical data inventory had similar criterion-related validity in each country cluster, it is important to examine the selection instrument for predictive bias or differential prediction across country clusters to evaluate its global utility (Society for Industrial and Organizational Psychology [SIOP], 2003). Predictive bias examines if the slope and/or intercepts of the “regression line relating the predictor (i.e., biographical data inventory) to the criterion (i.e., job performance) are different for one group than for another” (SIOP, 2003, p. 32). In the context of the current research, this provides a stricter test of the equivalence of prediction between country clusters compared to the omnibus χ^2 test, as it looks for not only slope but also intercept differences (Aguinis, Culpepper, & Pierce, 2010; Cleary, 1968; Lautenschlager & Mendoza, 1986).

This procedure involves testing, via ΔR^2 , a series of regression models that include various combinations of the predictor (i.e., biographical data inventory), subgroup (i.e., country cluster membership), and predictor-subgroup interaction. The first comparison is between a regression model relating scores on the predictor to the criterion (Model 1) versus a model with all three terms (Model 2). If there is a significant increase in the proportion of variance explained

by including the dummy coded group membership variable and cross product term, then there is evidence of differential prediction. This suggests that a common regression line is insufficient to account for the relationship between the predictor and criterion between two different groups. When this occurs, a series of additional models are compared to identify if the predictive bias is due to slope and/or intercept differences in the regression equation between majority (i.e., cross-validation U.S. sample) and minority (i.e., a country cluster) groups (Cleary, 1968; Lautenschlager & Mendoza, 1986).

It is important to note that in the high-stakes testing research literature, test bias has most commonly been examined with cognitive ability assessments. The consensus is that when differential prediction exists, it is usually in the form of intercept differences rather than slope differences (Aguinis et al., 2010). Practically speaking, when intercept differences are present, the use of a common regression equation results in the overprediction of scores for minority group members because the majority group has a larger intercept. Research examining the predictive bias of biographical data inventories is extremely limited. Rather, researchers have focused on standardized mean score subgroup differences, which tend to be smaller compared to cognitive ability assessments (Bobko & Roth, 2013; Bobko, Roth, & Potosky, 1999). Therefore, this suggests that it may be unlikely to observe intercept differences on a biographical data inventory between groups.

Despite the consensus in the research literature about the analysis approach and general findings regarding differential prediction, recent simulation research by Aguinis and colleagues (2010) has questioned these long held conclusions because of methodological and statistical artifacts (e.g., inadequate power, heterogeneity of error variance), which influence the accuracy of differential prediction analyses. Attenuated statistical power is a common problem that limits

researchers' ability to conclude if an assessment actually exhibits predictive bias. This happens because many test validation studies have small samples. As an example, Salgado (1998b) reported that from 1983 to 1994, the median sample size for personnel selection validation studies published in a variety of top tier journals (i.e., *Journal of Applied Psychology*, *Journal of Occupational and Organizational Psychology*, and *Personnel Psychology*) was 113. In addition, power issues in validation studies are magnified when testing differences in regression equations between majority and minority groups, especially when group sizes are unequal or when the homogeneity of within-subgroup error variance assumption is violated (Aguinis & Pierce, 1998; Aguinis, Petersen, & Pierce, 1999). Ultimately, this frequently results in inadequate power to detect slope and/or intercept differences between groups (Aguinis et al., 2010). Therefore, it is plausible that for underpowered or problematic studies, a finding of no differential prediction is actually a Type II error (i.e., failing to reject the hypothesis that there is a moderating effect in the population). It is worth noting that these issues are not unique to differential prediction analyses but are broadly common when using moderated multiple regression or conducting test validation research in organizational settings (Aguinis et al., 2010).

Given these recent advancements in differential prediction research, prior to running the step-down hierarchical regression procedure outlined by Lautenschlager and Mendoza (1986), I tested two critical assumptions. Namely, if there was adequate power to detect the predictor-subgroup moderating effect and if there was homogeneity of error variance across each country cluster. It is important to check and meet both of these assumptions (or take measures to rectify the issue[s]) to avoid decision errors and ensure accurate conclusions (Aguinis & Pierce, 1998).

I used the procedure and program (POWER) outlined by Aguinis, Pierce, and Stone-Romero (1994) to estimate the statistical power to detect the predictor-subgroup interaction for

each analysis comparing the majority group (i.e., U.S. cross-validation) and minority group (i.e., a single country cluster). Using inputs from Table 3 (i.e., sample size in each group, sample-based correlation between the predictor and criterion), the estimated statistical power to detect the predictor-subgroup interaction for each hierarchical regression analysis was 1.0. This indicated that, due to the relatively small difference in the magnitude of the correlation coefficient between each country cluster and the moderately large sample sizes in each group, there was adequate power to detect predictor-subgroup interactions for each comparison. That is, there is greater confidence that null findings regarding differential prediction for the biographical data inventory are likely be accurate rather than Type II errors.

Next, homogeneity of error variance was examined using the procedure and program (ALTMMR) outlined by Aguinis, Petersen, and Pierce (1999). Inputs from Table 2 (i.e., mean and standard deviation for the predictor and criterion in each sub-group) and Table 3 (i.e., sample size and validity in each sub-group), were used to calculate DeShon and Alexander's (1996) 1.5 rule of thumb⁷ and Barlett's M to test for homogeneity of error variance. If the homogeneity of error variance assumption is violated, ALTMMR provides two alternative indices, James's (1951) second-order approximation (J), and Alexander's (1994) normalized t approximation (A) (Alexander & Govern, 1994). These alternative statistics provide comparable results to the traditional F test (discussed below) when this assumption is violated to examine the significance of the predictor-subgroup interaction to determine if an assessment exhibits predictive bias.

Results for the homogeneity of error variance tests between country clusters are presented in Table 4. First, the omnibus test of homogeneity of error variance indicated that the assumption was not met for at least one country cluster. Follow-up individual comparisons between the

⁷ Based on extensive Monte Carlo simulations, DeShon and Alexander (1996) found that the power of moderated multiple regression is not impacted until the error variance of one group is approximately 1.5 times greater than the error variance of the other subgroup.

majority group (U.S. cross-validation) and minority group (a single country cluster) indicated that the homogeneity of error variance assumption was met when comparing the U.S. cross-validation sample to the Confucian Asia, Latin Europe, and Southern Asia country clusters. The results for the Anglo-Saxon and Germanic clusters indicated that one (Bartlett's M) or both of the tests for equal error variances was significant, respectively. When a discrepancy such as that observed in the Anglo-Saxon cluster occurs, Aguinis and colleagues (1999) recommend examining James' J , Alexander's A , and the traditional F test to see if these statistics provide a consistent conclusion. These additional tests indicated that, while there was heterogeneity of error variance between the U.S. cross-validation sample and the Anglo-Saxon country cluster, the predictor-subgroup interaction was non-significant according to the two alternative indices and the traditional F test (these results are discussed below and presented in Table 5). For the Germanic cluster, both tests for homogeneity of error variance were significant. Therefore, rather than looking at the traditional F test results for the predictor-subgroup interaction, it is more appropriate to examine James' J and Alexander's A , which account for this heterogeneity. These indices indicated that the predictor-subgroup interaction was non-significant. Overall, this analysis indicated that the homogeneity of error variance assumption was met for the majority of the country clusters (i.e., Confucian Asia, Latin Europe, and Southern Asia). When combined with the results of the power analysis, there were few concerns for clusters that exhibited heterogeneity of error variance (i.e., Anglo-Saxon and Germanic).

The results of the power analysis and homogeneity of error variance tests for each comparison indicated that it was appropriate to use Lautenschlager and Mendoza's (1986) aforementioned step-down hierarchical multiple regression approach to examine if the biographical data inventory exhibited differential prediction between the majority group (i.e.,

U.S. cross-validation sample) and each country cluster. The results of these analyses are presented in Table 5.

This table shows, for each comparison cluster, the proportion of variance explained by Model 1 (R^2), the increase (Δ) in R^2 by including the dummy coded group membership variable and cross-product term (i.e., Model 2), and the significance test of this change (distributed as an F test). If this test is statistically significant, then there is evidence of differential prediction. When this occurs, it is necessary to compare additional regression models to identify if the bias is due to slope and/or intercept differences (Cleary, 1968; Lautenschlager & Mendoza, 1986). Results indicated that the increase in the proportion of variance explained from Model 1 to Model 2 was non-significant for each comparison cluster, $F_{\text{Anglo-Saxon}}(2, 866) = 0.42, ns, \Delta R^2 = 0.001$; $F_{\text{Confucian Asia}}(2, 959) = 0.76, ns, \Delta R^2 = 0.001$; $F_{\text{Germanic}}(2, 692) = 1.40, \Delta R^2 = 0.004, ns$; $F_{\text{Latin Europe}}(2, 703) = 1.01, ns, \Delta R^2 = 0.003$; and $F_{\text{Southern Asia}}(2, 617) = 0.86, ns, \Delta R^2 = 0.002$. That is, evidence suggests that the U.S. developed biographical data inventory did not exhibit differential prediction in any country cluster. Therefore, a common regression line could account for the relationship between the predictor and criterion for each cluster and it was not necessary to conduct follow-up model comparisons. This provides additional evidence that the biographical data inventory was equally effective when transported to each country cluster.

In summary, results indicated that the biographical data inventory was a significant predictor of managerial job performance in the U.S. development sample and remained a strong predictor (with minimal shrinkage) of performance in the U.S. cross-validation sample. When the same U.S. scoring key was applied to each of the five country clusters, it remained a significant predictor of job performance. An omnibus test of the equality of independent correlations indicated that the magnitude of the validity coefficient in the U.S. cross-validation

sample was not significantly different from the magnitude observed in each of the other country clusters. Lastly, step-down hierarchical regression analyses indicated that the biographical data inventory did not exhibit differential prediction in any country cluster. Collectively, these analyses provide evidence to support my hypothesis that empirically keyed worker-oriented biographical data inventories, which use rigorous job analysis methods and theory as the foundation for content generation, generalize to managers in each of the country clusters examined.

Reproducibility, however, is a hallmark of good science and scientific progress (Popper, 1959; Open Science Collaboration, 2012). Therefore, Study 2 seeks to further test these preliminary conclusions regarding the cross-national generalizability of biographical data with a new biographical data assessment, which uses the same development approach, in a unique manufacturing (vs. managerial) sample of job incumbents from overlapping (Southern Asia, Confucian Asia, and Germanic) and unique (Eastern Europe, Latin America) country clusters. If replicated, this will provide additional support for the hypothesis that rigorously developed biographical data inventories will broadly generalize to other countries.

Table 3*Study 1 Biographical Data Inventory Criterion-Related Validity by Country Cluster*

Country Cluster	Countries	<i>N</i>	<i>r</i>	95% CI Lower	95% CI Upper
Anglo-Saxon	Canada, United Kingdom	328	0.35**	0.25	0.44
Confucian Asia	China, Japan	421	0.31**	0.22	0.39
Germanic	Belgium, Germany	154	0.29**	0.14	0.43
Latin Europe	France, Italy	165	0.26**	0.11	0.40
Southern Asia	India	79	0.33**	0.12	0.51
United States	Cross-validation	542	0.34**	0.26	0.41
United States	Development	1074	0.39**	0.35	0.45

Note. *N* = sample size; *r* = uncorrected validity coefficient obtained from applying the English empirical key; 95% CI = 95% confidence interval for *r*.

***p* < .01.

Table 4*Study 1 Homogeneity of Error Variance Between U.S. and Each Country Cluster*

Comparison Cluster	Error Variance Results		Alternative Moderation Tests	
	<i>D&A</i>	<i>M</i>	<i>J</i>	<i>A</i>
Omnibus Test	1:1.68*	18.27*	2.49	2.48
Anglo-Saxon	1:1.27	5.63*	0.01	0.01
Confucian Asia	1:1.10	1.13	N/A	N/A
Germanic	1:1.68*	14.56*	1.03	1.03
Latin Europe	1:1.29	3.82	N/A	N/A
Southern Asia	1:1.20	1.10	N/A	N/A

Note. Comparison Cluster = Cluster compared against U.S. cross-validation sample; *D&A* = DeShon and Alexander's (1996) 1.5 rule of thumb; *M* = Bartlett's *M*, *J* = James' *J*; *A* = Alexander's *A*; N/A = *J* and *A* are not reported when both tests indicated homogeneity of error variances because the *F* test is accurate.

* $p < .05$.

Table 5*Study 1 Differential Prediction Between U.S. and Each Country Cluster*

Comparison Cluster	R^2	ΔR^2	F	DF 1	DF 2	p
Anglo-Saxon	0.12	0.001	0.42	2	866	<i>ns</i>
Confucian Asia	0.11	0.001	0.76	2	959	<i>ns</i>
Germanic	0.11	0.004	1.40	2	692	<i>ns</i>
Latin Europe	0.11	0.003	1.01	2	703	<i>ns</i>
Southern Asia	0.11	0.002	0.86	2	617	<i>ns</i>

Note. Comparison Cluster = Cluster compared against U.S. cross-validation sample; R^2 = proportion of variance explained by model 1; ΔR^2 = change in R^2 from model 1 to model 2; F = significance test of the change in R^2 from model 1 to model 2; DF = Degrees of Freedom; p = significance associated with the F test.

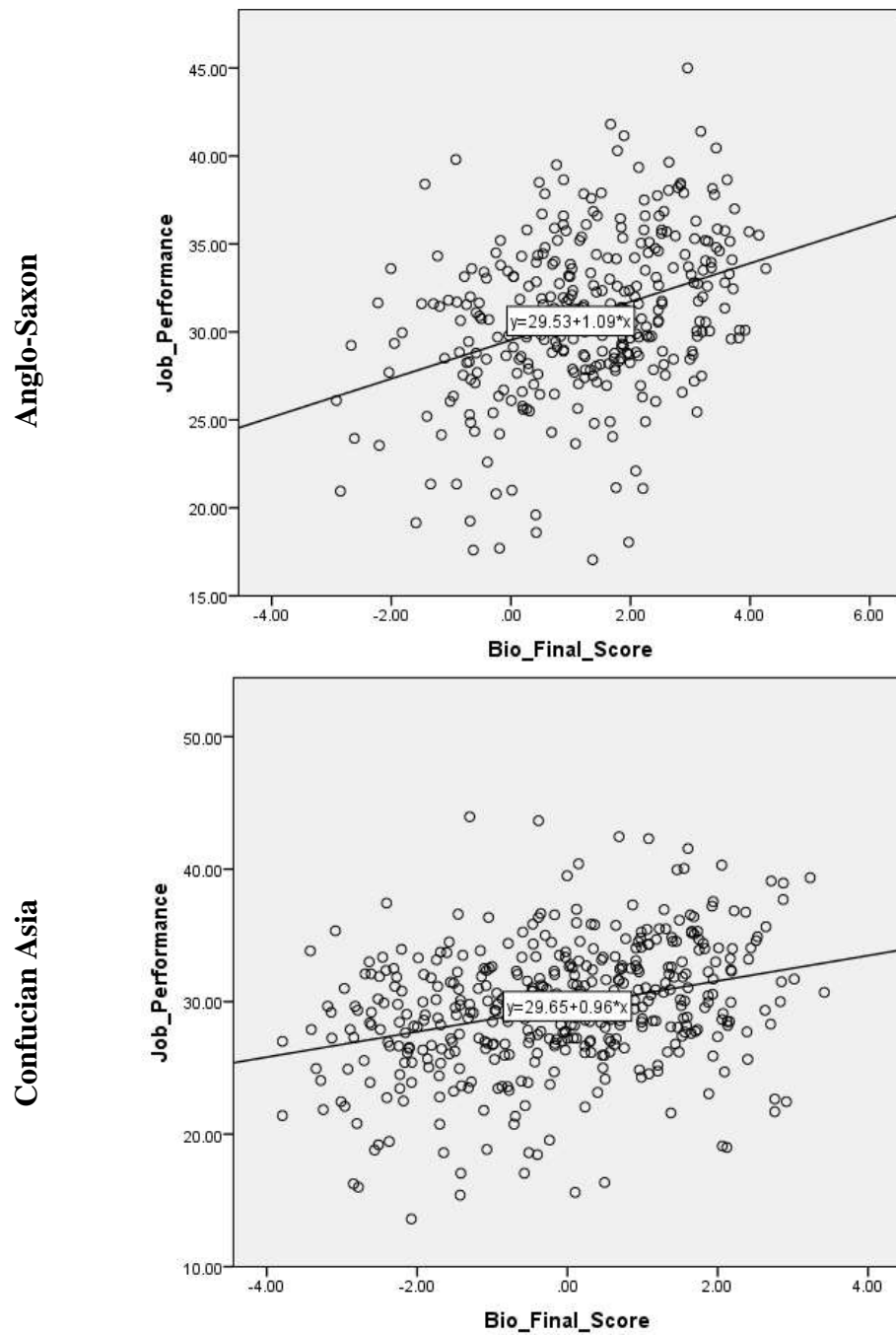


Figure 2. Study 1 Scatterplot of Criterion-Related Validity by Country Cluster.

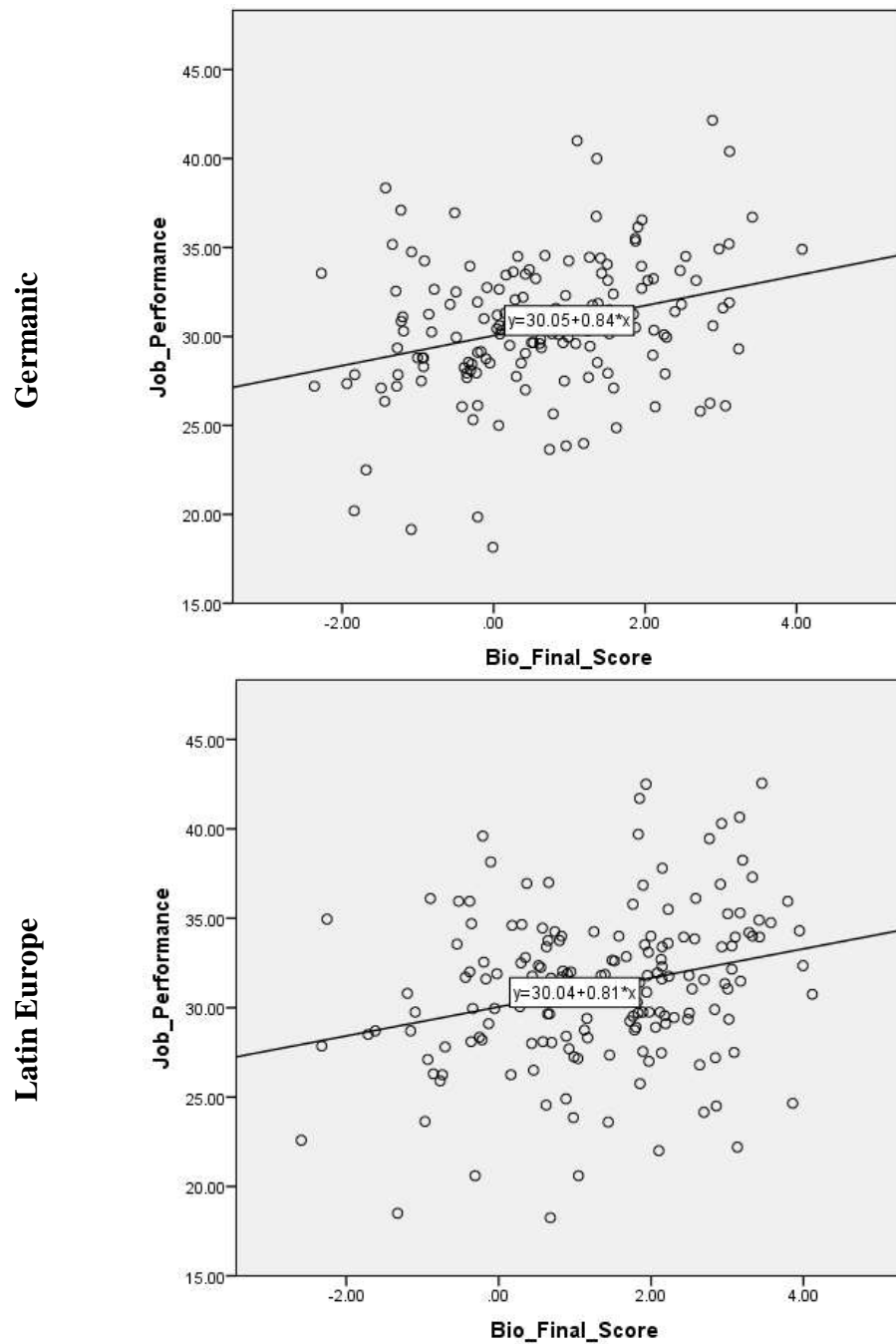


Figure 2. Study 1 Scatterplot of Criterion-Related Validity by Country Cluster (continued).

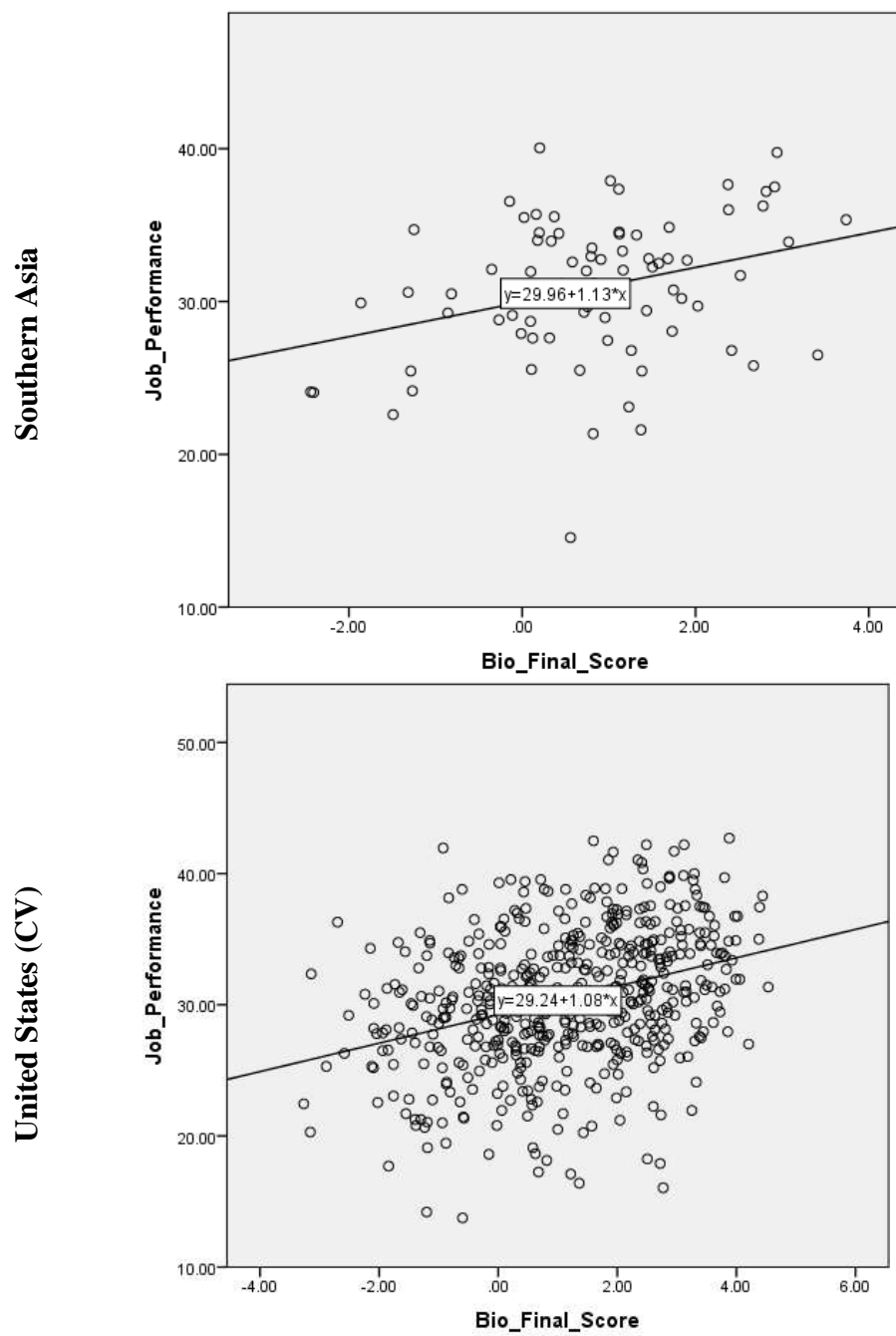


Figure 2. Study 1 Scatterplot of Criterion-Related Validity by Country Cluster (continued).

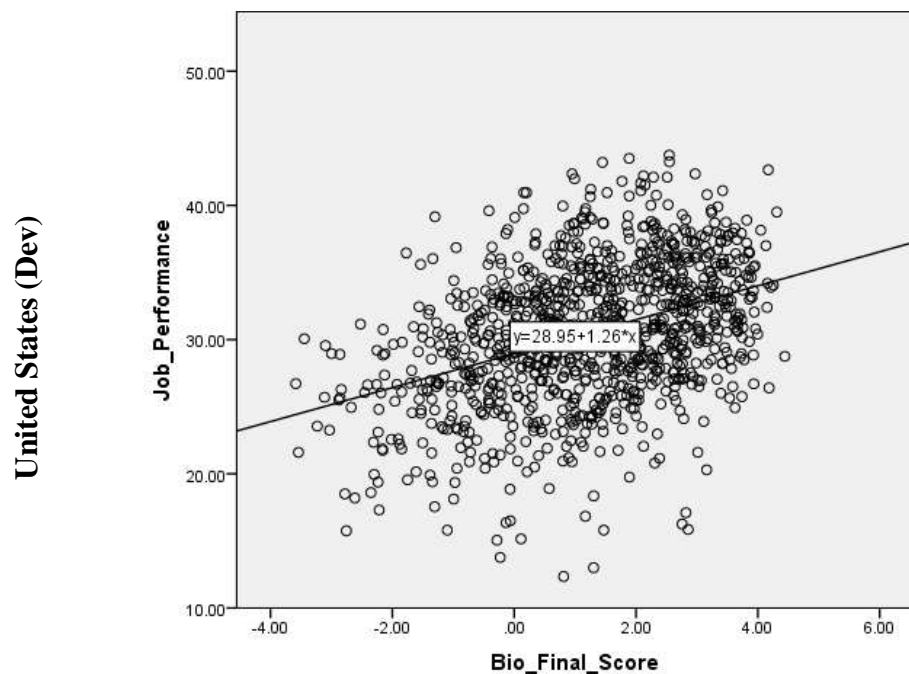


Figure 2. Study 1 Scatterplot of Criterion-Related Validity by Country Cluster (continued).

Method - Study 2

Participants

Participants were 1,210 job incumbents that worked for a single large multinational consumer goods company. All participants were manufacturing technicians and had a variety of different job titles (e.g., Manufacturing Production, Team Assembler, Chemical Plant and System Operator, Production Worker). The sample had more males (72.2%) compared to females (27.3%). The majority of the sample was from the United States ($N = 618$) but participants also came from Brazil ($N = 17$), China ($N = 189$), Germany ($N = 46$), Hungary ($N = 34$), India ($N = 142$), Japan ($N = 11$), Mexico ($N = 74$), Poland ($N = 36$), Russia ($N = 39$), and Singapore ($N = 4$). To increase statistical power, country clusters were created based on the GLOBE framework (House et al., 2004). This resulted in five clusters: Confucian Asia (China, Japan, Singapore), Eastern Europe (Hungary, Poland, Russia), German (Germany), Latin America (Mexico, Brazil), and Southern Asia (India).

As shown in Table 6, there were minor differences in the distribution of gender across each country cluster. These differences, however, ultimately had a little influence on the results. In addition, all participants had similar roles and responsibilities, regardless of where they worked. That is, the organization had many manufacturing plants around the world and each plant made similar products (e.g., laundry detergent, deodorant, shaving cream, toilet paper) in each local market. Therefore, the KSAOs needed to succeed in the various roles were believed to be similar – something that was confirmed through a global competency modeling project, which is discussed next.

Measures

Job performance.

Each employee's immediate supervisor assessed his or her performance with 14 items, which were organized into 10 performance competencies as identified through a separate manufacturing technician global competency modeling project. This project used interviews, focus groups, and questionnaires to identify these competencies and confirm that these critical dimensions were needed for performance across all manufacturing jobs (e.g., Manufacturing Production, Team Assembler). The resulting ten dimensions included KSAOs such as problem solving, collaboration, adaptability, execution, and safety. One to two items were used to assess each competency and each item was rated on a 7-point scale ranging from "Weak" to "Exceptional".

A confirmatory factor analysis, which specified a model with the item (for dimensions with a single item) or average (for dimensions with multiple items) score on each of the 10 competencies loading on a general performance factor, indicated adequate fit ($\chi^2(35) = 341.90$, RMSEA = 0.09, TLI = 0.94, SRMR = 0.03; Hu & Bentler, 1999). Coefficient alpha for the entire performance rating form indicated excellent internal consistency ($\alpha = .92$). Similar to the first study, these sources of statistical evidence suggested that there was also a general factor of job performance (Viswesvaran et al., 2005). Therefore, a composite performance criterion measure was created by averaging items within a competency and summing across each of the 10 competencies to create a unit-weighted (at the competency level) overall job performance score.

Biographical data inventory.

A 108-item biographical data inventory was administered to job incumbents as part of a concurrent validation study. The inventory items were developed to assess behaviors and

experiences associated with each of the 10 aforementioned dimensions (e.g., problem solving, collaboration, adaptability, execution, and safety) of manufacturing technician performance. Relevant psychological research and data gathered from the aforementioned global competency modeling project served as input when writing biographical data items. An example item written to assess the safety competency is, “What do you typically do when you see someone not following a recognized safety standard?”.

Items were reviewed by a panel of I-O psychologists and evaluated to ensure the content was globally appropriate (Mumford et al., 1996; Schaffer & Riordan, 2003). This item review was critical to ensure conceptual equivalence across country clusters since it was not possible to test for measurement invariance for the biographical data inventory because of statistical limitations (i.e., multidimensional data, non-normal multivariate distribution, insufficient N for some clusters). Therefore, the team of I-O psychologists evaluated items for: (1) relevance to the construct of interest, (2) contamination with other constructs, (3) social desirability, (4) bias regarding equal access, (5) potential for faking, (6) invasion of privacy, and (7) controllability of the behavior. Additionally, in the context of cross-national selection, items were reviewed to ensure that they were relevant across all country clusters and did not use colloquial language, idioms, or terms unique to one culture (Schaffer & Riordan, 2003).

An empirical scoring key for the biographical data inventory was developed and cross-validated in the sample of U.S. participants. Three-quarters of the sample ($N = 459$) was randomly selected as the development group and one-quarter ($N = 159$) was used to cross-validate the scoring key. The scoring key in study 2 was developed on a larger percentage of the U.S. sample (a priori) to improve statistical power, given that the total U.S. sample was slightly smaller compared to study 1. The biographical data inventory was empirically keyed with a point

biserial raw weight scoring approach. Items were retained if one or more response options had a statistically significant point biserial correlation with the composite criterion $\geq |.10|$ (Cucina et al., 2012). This cutoff was used because previous research has found this to be a useful cut point when determining what items to include in the final scoring key (Mumford & Owens, 1987). Including items with smaller relationships with the criterion has the potential to disproportionately increase error variance, which will impact the reliability and criterion-related validity of the instrument. This resulted in a final biographical data inventory with 35 items.

Procedure

A subset of manufacturing technicians from the multinational organization were invited to participate in the concurrent global validation study based on a stratified random sampling plan to ensure representativeness of the organization's manufacturing technician population by key demographic variables (e.g., gender, geography, and race [U.S. only]). Invited employees who elected to participate in the study completed the biographical data inventory in-person at local manufacturing sites. Participation was voluntary and confidential and demographic information for each participant was obtained from the organization's HRMS.

Willing participants completed the research form during a testing session in their native language. A professional translation agency completed all translations, which native speakers within the organization reviewed and compared to the English version. Any discrepancies were noted by the reviewers and resolved with the translator. Research participants from each country completed the inventory in the following language: United States = English, Brazil = Portuguese, China = Simplified Chinese, Germany = German, Hungary = Hungarian, India = Hindi, Japan = Japanese, Mexico = Spanish, Poland = Polish, Russia = Russian, and Singapore = English.

Concurrent with this data collection effort, research-only job performance ratings were gathered from each participant's immediate supervisor via an online survey. Prior to completing the job performance measure, supervisors received rater error and performance dimension training to reduce common rater biases and improve accuracy (Woehr & Huffcutt, 1994). In addition, the instructions for the rating form emphasized the confidential nature and research purposes of the ratings and instructed supervisors to rate honestly. Employees and supervisors completed all research forms during normal business hours and were compensated for their time with their standard salary

Table 6*Study 2 Demographic Information by Country Cluster*

Demographic Variable	<u>Overall</u>		<u>U.S. (Cross)</u>		<u>U.S. (Dev.)</u>		<u>Southern Asian</u>		<u>Eastern Europe</u>	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
<u>Gender</u>										
Male	874	72.2	99	62.3	310	67.5	140	98.6	81	74.3
Female	330	27.3	60	37.7	149	32.5	2	1.4	28	25.7
Prefer not to answer	6	0.5								
<u>Country</u>										
Brazil	17	1.4								
China	189	15.6								
Germany	46	3.8								
Hungary	34	2.8							34	31.2
India	142	11.7					142	100.0		
Japan	11	0.9								
Mexico	74	6.1								
Poland	36	3.0							36	33.0
Russia	39	3.2							39	35.8
Singapore	4	0.3								
United States	618	51.1	159	100.0	459	100.0				
<u>Race (Anglo-Saxon Only)</u>										
American Indian	3	0.5	1	0.6	2	0.4				
Asian	21	3.4	7	4.4	14	3.1				
Black	152	24.6	41	25.8	111	24.2				
Hispanic	52	8.4	12	7.5	40	8.7				
White	390	63.1	98	61.6	292	63.6				

Table 6 (continued)*Study 2 Demographic Information by Country Cluster*

Demographic Variable	<u>Confucian Asia</u>		<u>Germanic</u>		<u>Latin America</u>	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
<u>Gender</u>						
Male	128	62.8	41	89.1	75	82.4
Female	70	34.3	5	10.9	16	17.6
Prefer not to answer	6	2.9				
<u>Country</u>						
Brazil					17	18.7
China	189	92.6				
Germany			46	100.0		
Hungary						
India						
Japan	11	5.4				
Mexico					74	81.3
Poland						
Russia						
Singapore	4	2.0				
United States						

Results - Study 2

Descriptive Statistics

Table 7 provides descriptive statistics for manufacturing technicians on the biographical data inventory and job performance measure by country cluster. Supplemental analyses indicated that the mean score on the biographical data inventory is significantly different between country clusters, $F(5, 554.25) = 21.56, p < .001, \omega^2 = .05^8$. Relevant to the current research question, Tukey post hoc tests indicated that the mean score in the United States cross-validation sample ($M = 0.98$) is significantly greater than the mean score in the Confucian Asia ($M = -0.27; d = 0.76$), Eastern Europe ($M = 0.13; d = 0.49$), and Germanic ($M = 0.09; d = 0.55$) country clusters. The mean score on the job performance measure is not significantly different between the country clusters, $F(5, 745) = 1.20, ns, \omega^2 = .00$.

Given the relatively large sample size for many of the groups, it is important to emphasize that the effect size for the impact of country cluster on the biographical data inventory ($\omega^2 = .05$) is extremely small. This suggests that these statistically significant differences were of minor practical importance. Furthermore, some results are consistent with prior research about individuals from East Asia countries, who are more likely to use the midpoint of Likert scales compared to individuals from the U.S. (Chen et al., 1995). Additionally, since the score results for the selection instrument would be used to compare candidates for the same job, the majority of the people within the applicant pool are likely to be from the same country. Therefore, these

⁸ Levene's Test for Homogeneity of Variances was non-significant for both the biographical data inventory, $F(5,745) = 1.73, ns$, and the job performance measure, $F(5, 745) = 1.88, ns$. This indicated that it was appropriate to report the standard F statistic.

relatively minor mean score differences are less problematic from a practical standpoint because recruiters and hiring managers will rarely make between-country candidate comparisons.

The skewness and kurtosis values within each country cluster (Table 7) provide evidence that scores on the biographical data inventory and job performance measure are approximately normally distributed. All skewness and kurtosis values are less than 1.0. For country clusters with sample sizes less than 200, the formal statistical test of dividing the value by the standard error (distributed as z) is calculated. To preserve space, significant values are noted with an asterisk in Table 7. As shown in the table, only the biographical data inventory in the U.S. cross-validation sample exhibited slight negative skew. Visual inspection of the distribution for the biographical data inventory and job performance measure by country cluster (Figure 3) for manufacturing technicians provides (further) graphical evidence to support the relative normality of each variable by country cluster.

Coefficient alpha values for the job performance measure (Table 7) indicated generally excellent internal consistency reliability within each country cluster ($\alpha > .90$). The lowest observed value for coefficient alpha was in the Germanic country cluster ($\alpha = .86$), which also had the smallest sample size. Similar to Study 1, coefficient alpha was not calculated for the biographical data inventory because the empirical scoring key was multidimensional and measures of internal consistency are inappropriate for this assessment method (Mumford et al., 2012).

Hypothesis Testing

To evaluate the cross-national generalizability of the biographical data inventory, the U.S. manufacturing technician sample was randomly divided into development (3/4) and cross-validation (1/4) samples. After creating summary scores for the predictor and criterion measures

Table 7*Study 2 Descriptive Statistics by Country Cluster*

Country Cluster	Measure	<i>N</i>	<i>Mean (SD)</i>	<i>Skew (SE)</i>	<i>Kurtosis (SE)</i>	α
Confucian Asia	Biodata	204	-0.27 (1.53)	-0.05 (.17)	-0.74 (.34)	0.92
	Job Performance	204	46.34 (9.02)	-0.59 (.17)	0.78 (.34)	
Eastern Europe	Biodata	109	0.13 (1.73)	-0.22 (.23)	-0.75 (.46)	0.90
	Job Performance	109	43.73 (8.56)	-0.55 (.23)	0.71 (.46)	
Germanic	Biodata	46	0.09 (1.51)	0.12 (.35)	-0.25 (.69)	0.86
	Job Performance	46	45.73 (8.30)	-0.21 (.35)	-0.02 (.69)	
Latin America	Biodata	91	0.41 (1.48)	-0.31 (.25)	-0.68 (.50)	0.91
	Job Performance	91	45.92 (9.12)	0.00 (.25)	-0.79 (.50)	
Southern Asia	Biodata	142	0.78 (1.41)	-0.50 (.20)	-0.64 (.40)	0.94
	Job Performance	142	45.91 (10.35)	-0.25 (.20)	0.08 (.40)	
United States (CV)	Biodata	159	0.98 (1.74)	-0.56 (.19)*	-0.61 (.38)	0.94
	Job Performance	159	45.66 (9.39)	-0.20 (.19)	0.04 (.38)	
United States (Dev)	Biodata	459	0.91 (1.74)	-0.43 (.11)	-0.48 (.23)	0.95
	Job Performance	459	45.80 (10.05)	-0.35 (.11)	-0.13 (.23)	

Note. *N* = sample size; *SD* = standard deviation; *SE* = standard error; α = coefficient alpha; CV = cross-validation sample; Dev = development sample.

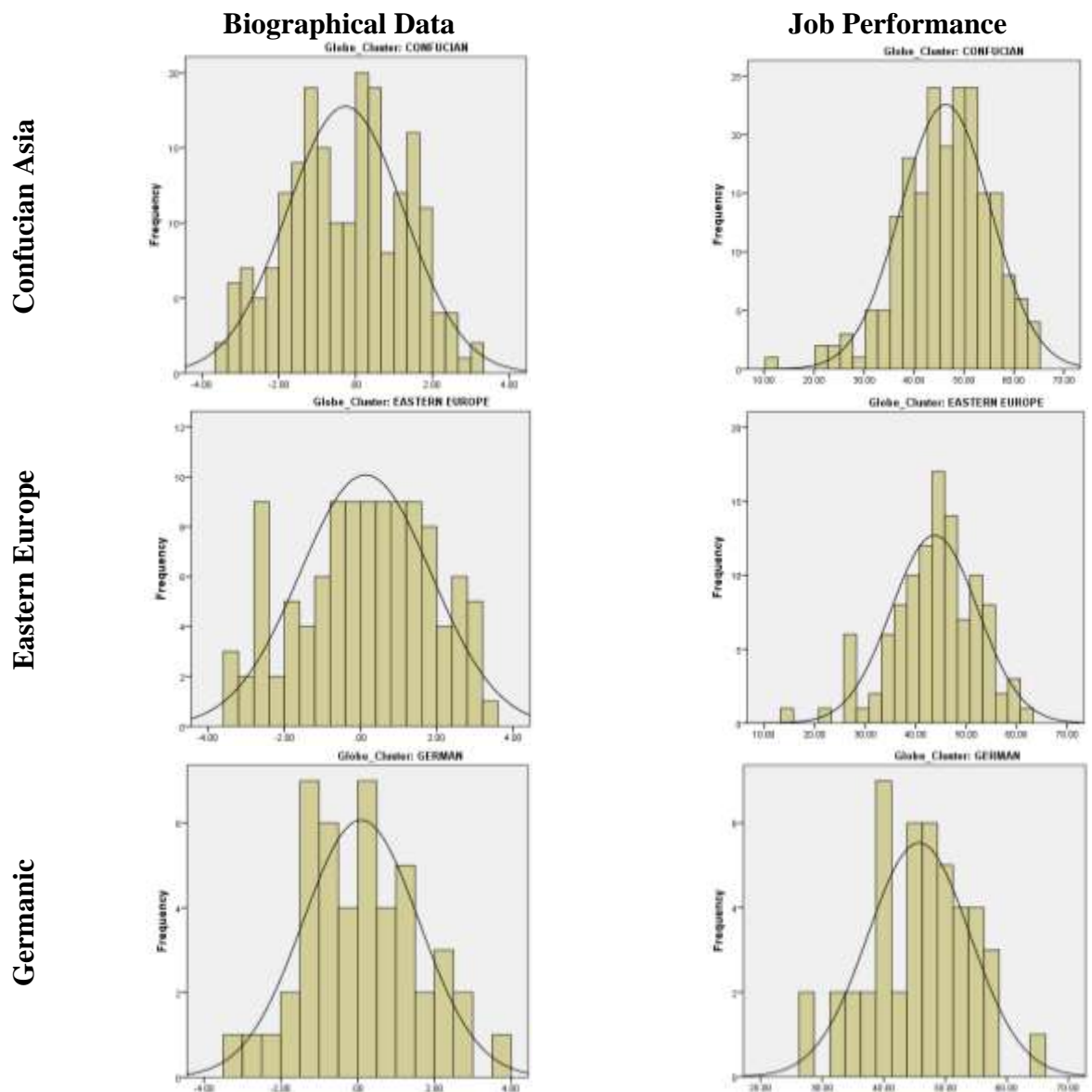
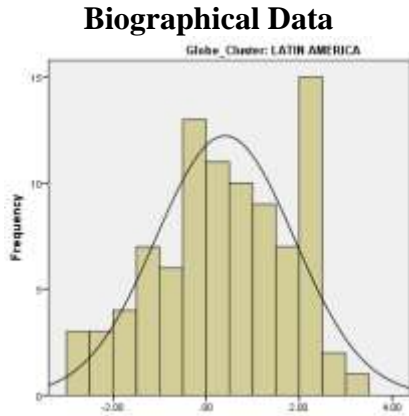
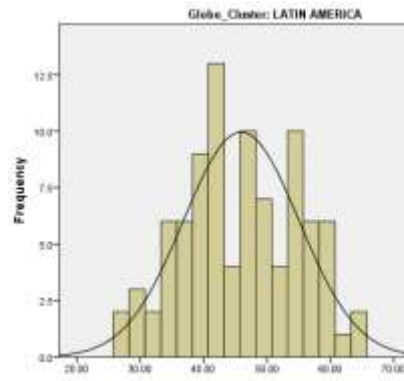


Figure 3. Study 2 Distribution of Variables by Country Cluster.

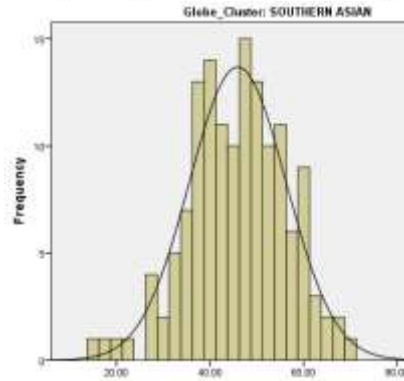
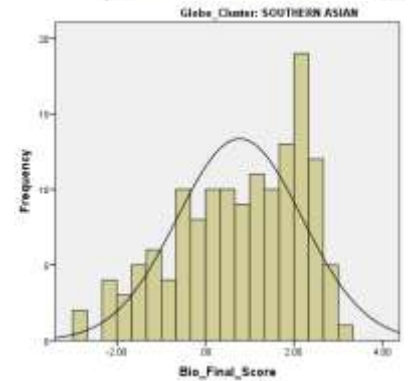
Latin America



Job Performance



Southern Asia



United States (CV)

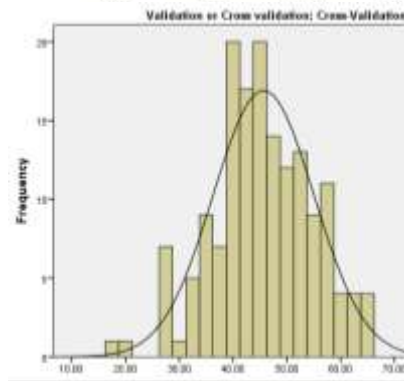
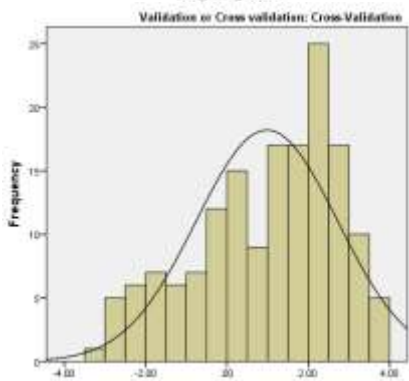


Figure 3. Study 2 Distribution of Variables by Country Cluster (continued).

United States (Dev)

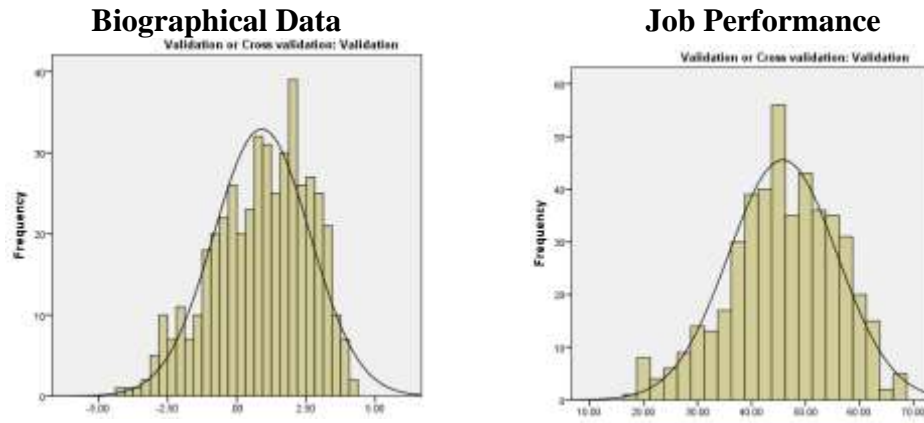


Figure 3. Study 2 Distribution of Variables by Country Cluster (continued).

according to the approaches outlined in the Method section⁹, results indicated that the biographical data inventory was significantly related to job performance in the U.S. development sample, $r(457) = 0.45, p < .01$.

To ensure that the empirical scoring key and its relationship with the criterion measure did not capitalize on chance or the unique characteristics of the development sample, the same scoring key was applied to the separate U.S. holdout sample. When the biographical data inventory's empirical scoring key was applied to this U.S. cross-validation sample, it remained a significant predictor of job performance, $r(157) = 0.33, p < .01$. The larger degree of shrinkage ($r = 0.12$) observed in the validity between the development ($r = 0.45$) and cross-validation ($r = 0.33$) sample in Study 2 is expected because the sample size in the development sample ($N = 459$) was smaller compared to Study 1 ($N = 1,074$). Empirical scoring keys based on smaller samples tend to show less stability (Cucina et al., 2012). In addition, the magnitude of the relationship and degree of shrinkage observed in the U.S. cross-validation managerial sample is consistent with cross-validated empirically keyed biographical data inventories reported in previous research (Beatty, 2013; Cucina et al., 2012; Schmidt & Hunter, 1998).

Next, the same empirical scoring key was separately applied to respondents from each of the five country clusters to examine the validity of the U.S. empirical scoring key in each region. Given that some country clusters had relatively small sample sizes (i.e., Germanic, Latin America), it is important to reiterate that the minimum sample size is 49 to have power equal .80 when $\rho = .35$ with a 1-tailed test (Faul et al., 2009). Only the Germanic country cluster ($N = 46$) had a sample size below this minimum recommended value. However, given that the purpose of

⁹ The manufacturing technician biographical data inventory was empirically keyed and scored with a point biserial raw weight scoring approach. Items included in the final key had a response option that had a statistically significant point biserial correlation with the composite criterion $\geq |.10|$.

Study 2 was to replicate the results of Study 1, it was included in the analyses. All other country clusters had a sufficient sample size to achieve adequate power.

Results indicated that the inventory was significantly related to job performance in each of the clusters: (a) Confucian Asia, $r(202) = 0.34, p < .01$, (b) Eastern Europe, $r(107) = 0.29, p < .01$, (c) Germanic, $r(44) = 0.34, p < .05$, (d) Latin America, $r(89) = 0.26, p < .05$, and (e) Southern Asia, $r(140) = 0.32, p < .01$. In addition, the magnitude of the validity coefficient observed in each country cluster appeared to be similar to the validity observed in the U.S. cross-validation sample ($r = 0.33$). Collectively, these results provide preliminary additional support for the cross-national generalizability of a biographical data inventory in a sample of manufacturing technicians because summary scores on the biographical data inventory, which were derived from the same U.S.-based empirical scoring key, were significantly related to job performance in the U.S. cross-validation sample and each country cluster. Refer to Table 8 and Figure 4 for detailed information about the criterion-related validity of the biographical data inventory by country cluster for manufacturing technicians.

Then, an omnibus test of the equality of independent correlations was used to examine the hypothesis that the magnitude of the observed validity in the U.S. cross-validation sample was not significantly different from the magnitude in each of the other country clusters (Chen & Popovich, 2002). This test is important to show that not only does the biographical data inventory significantly relate to job performance in each country cluster but that the strength of the relationship is similar in each region. Results indicated that there were no significant differences between the validity coefficients in the manufacturing sample by country cluster, $\chi^2(5) = 0.63, ns$. That is, the magnitude of the validity coefficient in each sample was not significantly different from each other. This omnibus test provides further evidence to support

the cross-national generalizability of the biographical data inventory, developed in the U.S., to a diverse sample of Confucian Asia (China, Japan, Singapore), Eastern Europe (Hungary, Poland, Russia), Germanic (Germany), Latin America (Mexico, Brazil), and Southern Asia (India) country clusters.

After finding that the biographical data inventory had similar criterion-related validity in each country cluster, I examined the manufacturing technician selection instrument for predictive bias across country clusters to evaluate its global utility (Society for Industrial and Organizational Psychology [SIOP], 2003). This provides a stricter test of the equivalence of prediction between country clusters compared to the omnibus χ^2 test, as it looks for not only slope but also intercept differences (Aguinis, Culpepper, & Pierce, 2010; Cleary, 1968; Lautenschlager & Mendoza, 1986).

Prior to running the step-down hierarchical regression procedure outlined by Lautenschlager and Mendoza (1986), I checked if there was adequate power to detect the predictor-subgroup moderating effect and homogeneity of error variance across each country cluster. It is important to check both of these pieces of information to avoid decision errors and ensure accurate conclusions (Aguinis & Pierce, 1998).

Preliminary power analyses using the program POWER showed that the ability to detect a predictor-subgroup interaction, with the U.S. cross-validation sample as the majority group, were limited for some comparisons such as the Germanic (Power = .33) and Latin America (Power = .43) country clusters (Aguinis et al., 1994). Therefore, the U.S. development sample was combined with the U.S. cross-validation sample to create the majority group for all differential prediction analyses. After combining these samples, inputs from Table 8 (i.e., sample size in each group, sample-based correlation between the predictor and criterion) indicated that

the estimated statistical power to detect the predictor-subgroup interaction for each hierarchical regression analysis comparison was 1.0. Therefore, combining the two U.S. samples enabled greater confidence that null findings regarding differential prediction for the biographical data inventory would likely be accurate rather than Type II errors. Additionally, since the criterion-related validity was fairly similar in the U.S. development and cross-validation samples, the benefit of increased statistical power outweighed the trade-offs of this post hoc decision.

Next, homogeneity of error variance was examined using the procedure and program (ALTMMR) outlined by Aguinis, Petersen, and Pierce (1999). Inputs from Table 7 (i.e., mean and standard deviation for the predictor and criterion in each sub-group) and Table 8 (i.e., sample size and validity in each sub-group), were used to calculate DeShon and Alexander's (1996) 1.5 rule of thumb and Barlett's M to test for homogeneity of error variance. If the homogeneity of error variance assumption was violated, James's (1951) second-order approximation (J) and Alexander's (1994) normalized t approximation (A) (Alexander & Govern, 1994) were examined to evaluate the significance of the predictor-subgroup interaction.

Results testing for homogeneity of error variance between country clusters are presented in Table 9. First, the two statistics for the omnibus test of homogeneity of error variance provided discrepant results. The DeShon and Alexander (1996) 1.5 rule of thumb indicated that the ratio of error variance (1:1.58) was slightly larger than the cutoff value of 1:1.50. This suggests that power may be impacted for an omnibus differential prediction analysis. Bartlett's M , however, was non-significant. Based on these results, follow-up individual comparisons between the majority group (U.S. cross-validation) and minority group (a single country cluster) indicated that the homogeneity of error variance assumption was met when comparing the U.S. cross-validation sample to the Confucian Asia, Eastern Europe, Germanic, Latin America, and

Southern Asia country clusters. Therefore, this analysis indicated that while the homogeneity of error variance assumption was not met when examining all country clusters simultaneously, it was met individually for each country cluster when compared the total U.S. manufacturing technician sample.

The results of the power analysis and homogeneity of error variance tests indicated that use of Lautenschlager and Mendoza's (1986) aforementioned step-down hierarchical multiple regression approach would have sufficient accuracy and not be susceptible to common methodological issues (Aguinis et al., 2010). The results of these analyses are presented in Table 10. This table shows, for each comparison cluster, the proportion of variance explained by Model 1 (R^2), the increase (Δ) in R^2 by including the dummy coded group membership variable and cross-product term (i.e., Model 2), and the significance test of this change (distributed as an F test). If this test was statistically significant, then there is evidence of differential prediction. When this occurs, additional regression models are compared to identify if the bias is due to slope and/or intercept differences (Cleary, 1968; Lautenschlager & Mendoza, 1986).

Results indicated that the increase in the proportion of variance explained from Model 1 to Model 2 was non-significant for the all country clusters, $F_{\text{Confucian Asia}}(2, 818) = 0.77, ns, \Delta R^2 = .001$; $F_{\text{Eastern Europe}}(2, 723) = 1.57, ns, \Delta R^2 = .004$; $F_{\text{Germanic}}(2, 660) = 1.18, ns, \Delta R^2 = .003$; $F_{\text{Latin America}}(2, 705) = 1.65, ns, \Delta R^2 = .004$; $F_{\text{Southern Asia}}(2, 756) = 0.18, ns, \Delta R^2 = .000$. That is, evidence suggests that the U.S. developed biographical data inventory did not exhibit differential prediction in any country cluster. Therefore, a common regression line could account for the relationship between the predictor and criterion for each cluster and it was not necessary to conduct follow-up model comparisons. This provides additional evidence that the manufacturing

technician biographical data inventory was equally effective when transported to each of these country clusters. Refer to Table 10 for the results of all differential prediction analysis.

In summary, results indicated that the biographical data inventory was a significant predictor of manufacturing technician job performance in the U.S. development sample and remained a strong predictor (with minimal shrinkage) of performance in the U.S. cross-validation sample. When the same scoring key was applied to each of the five country clusters, it remained a significant predictor of job performance. An omnibus test of the equality of independent correlations indicated that the magnitude of the validity coefficient in the U.S. cross-validation sample was not significantly different from the magnitude observed in each of the other country clusters. Lastly, step-down hierarchical regression analyses indicated that the biographical data inventory did not exhibit differential prediction in any country cluster.

Collectively, the results of Study 2 serve to replicate the findings of Study 1 but in a separate occupational group and with additional country clusters. These results provide additional evidence to support my hypothesis that empirically keyed worker-oriented biographical data inventories, which use rigorous job analysis methods and theory as the foundation for content generation, generalize to employees in a broad sample of country clusters.

Table 8*Study 2 Biographical Data Inventory Criterion-Related Validity by Country Cluster*

Country Cluster	Countries	<i>N</i>	<i>r</i>	95% CI Lower	95% CI Upper
Confucian Asia	China, Japan, Singapore	204	0.34**	0.21	0.45
Eastern Europe	Hungary, Poland, Russia	109	0.29**	0.11	0.46
Germanic	Germany	46	0.34*	0.05	0.57
Latin America	Mexico, Brazil	91	0.26*	0.05	0.44
Southern Asia	India	142	0.32**	0.16	0.46
United States	Cross-validation	159	0.33**	0.19	0.46
United States	Development	459	0.45**	0.37	0.52

Note. *N* = sample size; *r* = uncorrected validity coefficient obtained from applying the English empirical key; 95% CI = 95% confidence interval for *r*.

**p* < .05.

***p* < .01.

Table 9*Study 2 Homogeneity of Error Variance between U.S. and Each Country Cluster*

Comparison Cluster	Error Variance Results		Alternative Moderation Tests	
	<i>D&A</i>	<i>M</i>	<i>J</i>	<i>A</i>
Omnibus Test	1:1.58*	6.63	4.86	4.81
Confucian Asia	1:1.12	0.92	N/A	N/A
Eastern Europe	1:1.20	1.44	N/A	N/A
Germanic	1:1.32	1.49	N/A	N/A
Latin America	1:1.04	0.05	N/A	N/A
Southern Asia	1:1.20	1.91	N/A	N/A

Note. Comparison Cluster = Cluster compared against U.S. sample; *D&A* = DeShon and Alexander's (1996) 1.5 rule of thumb; *M* = Bartlett's *M*, *J* = James' *J*; *A* = Alexander's *A*; N/A = *J* and *A* are not reported when both tests indicated homogeneity of error variances because the *F* test is accurate.

* $p < .05$.

Table 10*Study 2 Differential Prediction between U.S. and Each Country Cluster*

Comparison Cluster	R^2	ΔR^2	F	DF 1	DF 2	p
Confucian Asia	0.16	0.001	0.77	2	818	<i>ns</i>
Eastern Europe	0.17	0.004	1.57	2	723	<i>ns</i>
Germanic	0.17	0.003	1.18	2	660	<i>ns</i>
Latin America	0.16	0.004	1.65	2	705	<i>ns</i>
Southern Asia	0.16	0.000	0.18	2	756	<i>ns</i>

Note. Comparison Cluster = Cluster compared against U.S. sample; R^2 = proportion of variance explained by model 1; ΔR^2 = change in R^2 from model 1 to model 2; F = significance test of the change in R^2 from model 1 to model 2; DF = Degrees of Freedom; p = significance associated with the F test.

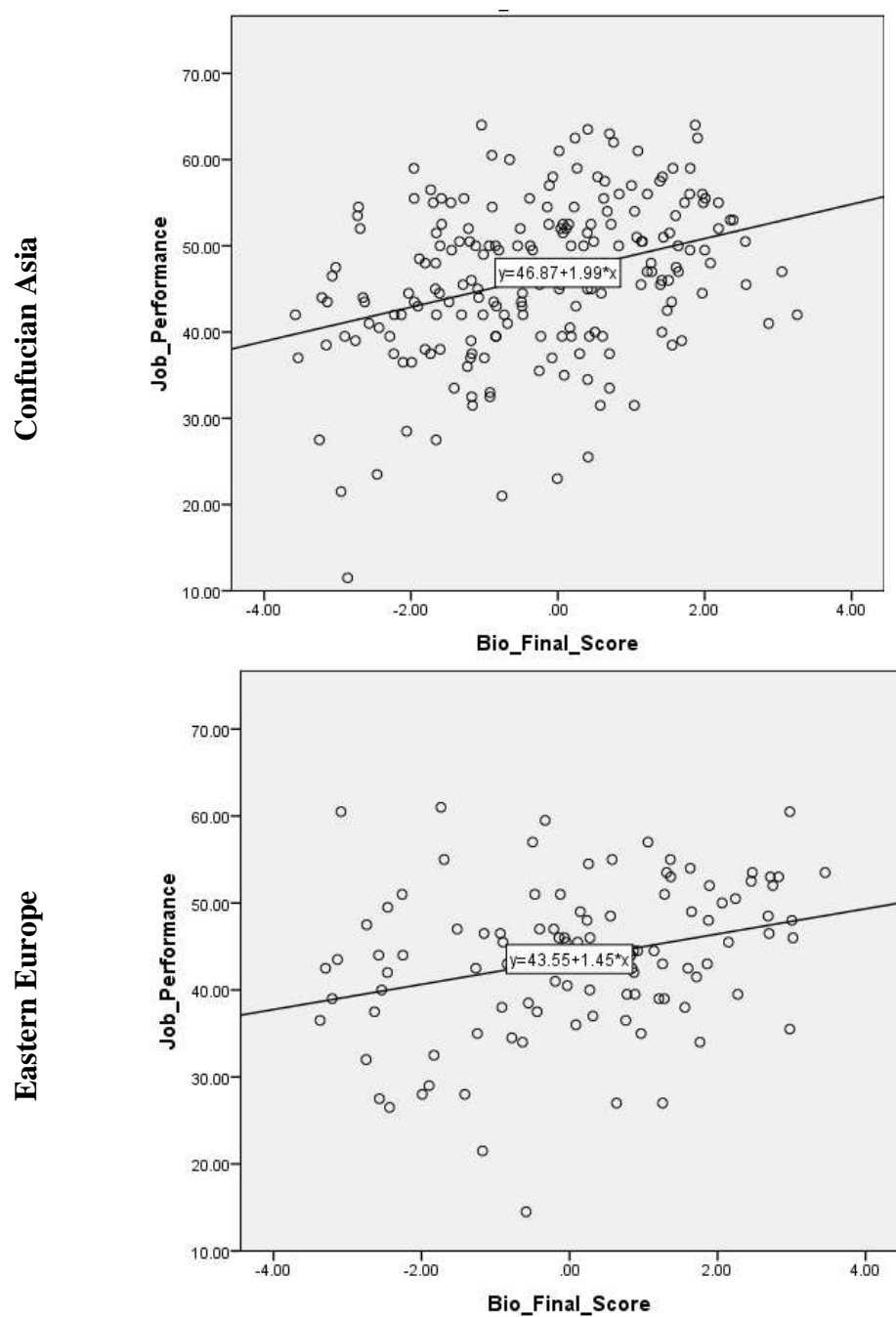


Figure 4. Study 2 Scatterplot of Criterion-Related Validity by Country Cluster.

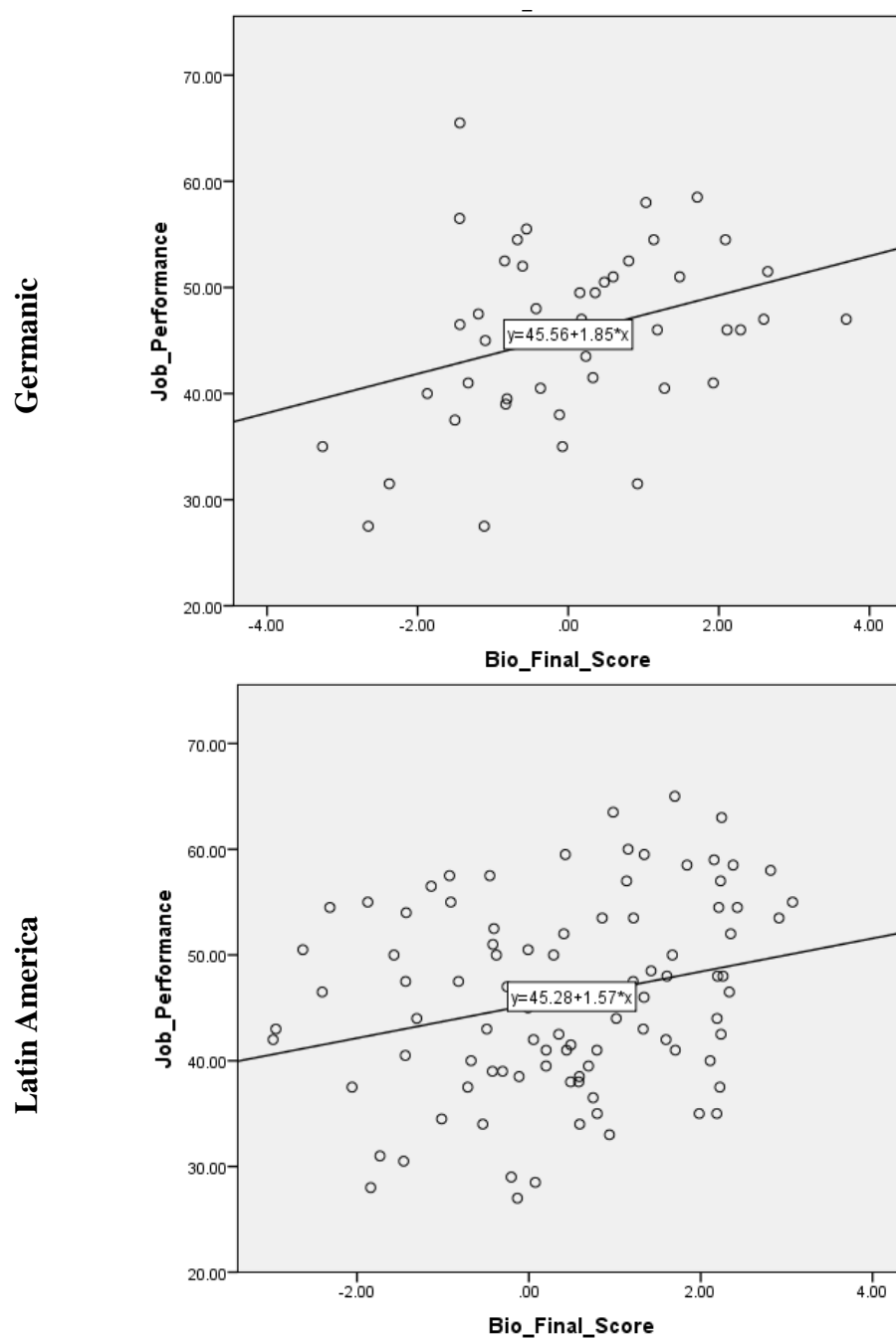


Figure 4. Study 2 Scatterplot of Criterion-Related Validity by Country Cluster (continued).

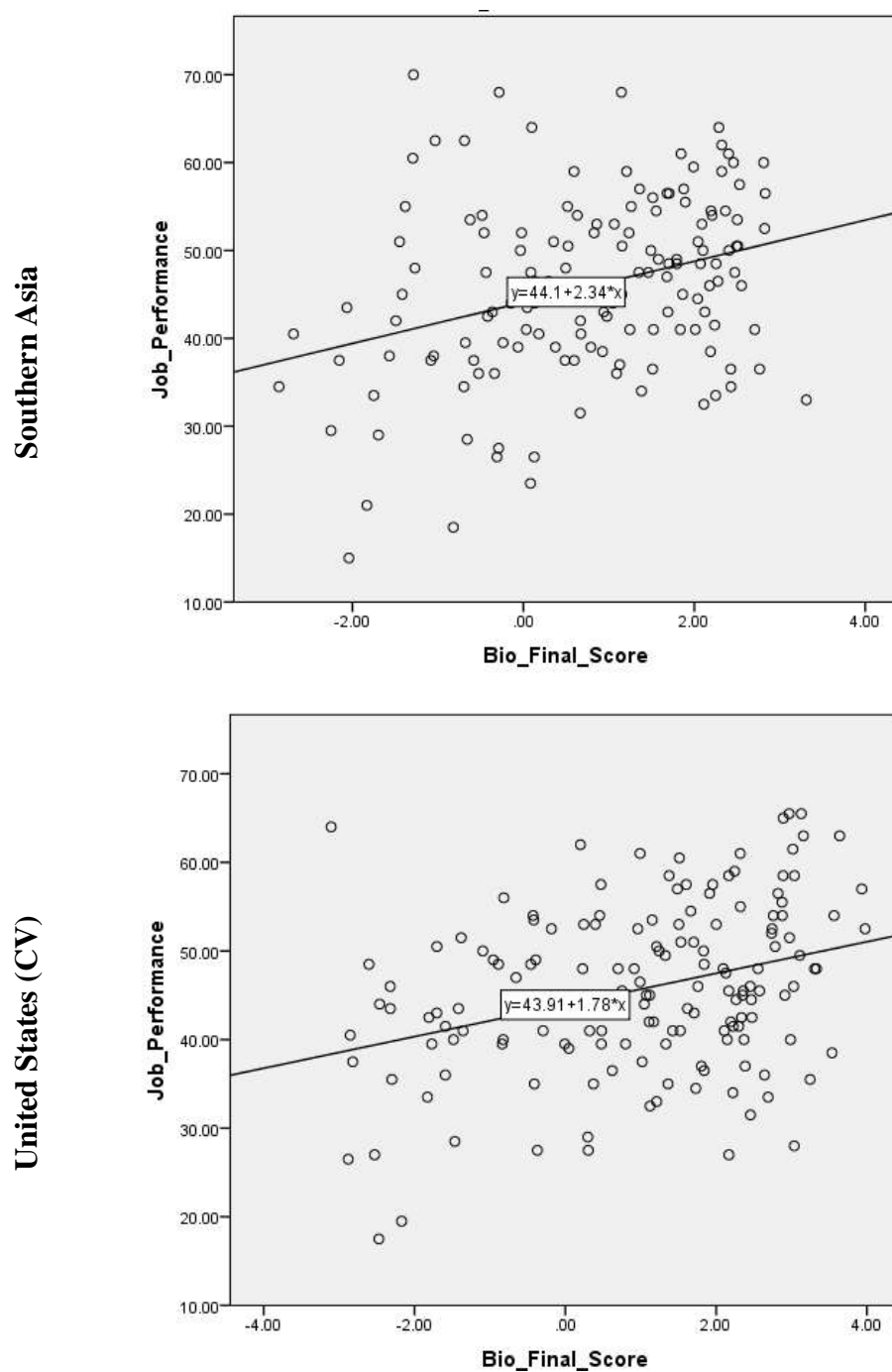


Figure 4. Study 2 Scatterplot of Criterion-Related Validity by Country Cluster (continued).

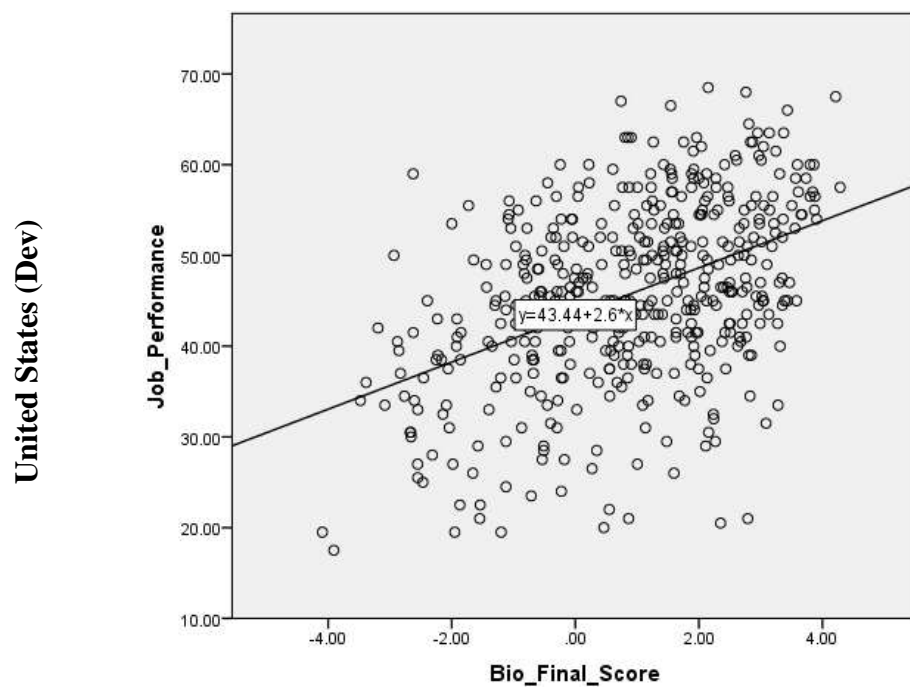


Figure 4. Study 2 Scatterplot of Criterion-Related Validity by Country Cluster (continued).

Discussion

Personnel selection is a key function for human resources departments. It is concerned with identifying the individuals who will make up an organization's workforce. In an increasingly global economy, multinational organizations must consider how to design and implement effective global selection systems. The present study extends current knowledge about the effectiveness of various selection tools when transported to other countries or cultures. Specifically, results suggest that biographical data inventories, which use assessment development best practices, have generalized validity across a diverse sample of country clusters.

Two studies involving participants from 7 country clusters, across four continents, and two different occupational groups (Study 1: managers; Study 2: manufacturing technicians), found evidence to support the hypothesis that empirically keyed worker-oriented biographical data inventories, which use rigorous job analysis methods and theory as the foundation for content generation, generalize to all country clusters examined. Results for Study 1 and Study 2 indicated that the biographical data inventories were significant predictors of job performance in the U.S. development samples and remained strong predictors (with minimal shrinkage) of performance in the U.S. cross-validation samples. When the same scoring key was applied to each of the five country clusters in Study 1 (Anglo-Saxon, Confucian Asia, Germanic, Latin Europe, and Southern Asia) and Study 2 (Confucian Asia, Eastern Europe, Germanic, Latin America, and Southern Asia), the biographical data inventories remained significant predictors of job performance for both occupational groups. Omnibus tests of the equality of independent correlations indicated that the magnitude of the validity coefficients in the U.S. cross-validation

samples were not significantly different from the magnitude observed in each of the other country clusters. Lastly, step-down hierarchical regression analyses indicated that the biographical data inventories did not exhibit differential prediction in any country cluster for Study 1 or Study 2. It is worth acknowledging that the results, especially the magnitude of the validity coefficients, were remarkably similar across country cluster. These findings may have occurred because of a variety of factors unique to this single multinational organization (e.g., comprehensive global competency modeling projects, extensive rater training, support from managers and leadership to have adequate time to complete research materials), which increased the rigor of the research. I expand on this point when discussing study limitations.

It is important to note that the biographical data inventory exhibited small standardized mean score differences for some country clusters when compared to the U.S. cross-validation sample. In Study 1, the mean score in the U.S. cross-validation sample was significantly greater than the mean score in the Germanic and Confucian Asia country clusters. In Study 2, the mean in the U.S. cross-validation sample was significantly greater than the mean score in the Confucian Asia, Eastern Europe, and Germanic country clusters. Using standard interpretation guidelines, the effect size for the overall test for these mean score differences was below the threshold for a small effect size. Therefore, it suggests that these findings are likely to have minor practical importance, especially when an applicant pool for a job is only comprised of individuals from the same country. In this situation, these small mean score differences are unlikely to impact who is ultimately selected. Organizations, however, should be aware of these differences when comparing candidates from different country clusters (e.g., internal promotion opportunities, global development programs) because some of the specific *d*-values between country clusters were fairly large and may influence HR decisions or outcomes (Bobko & Roth,

2013; Bobko, Roth, & Potosky, 1999). It is important to be aware of these differences if directly comparing candidates across geographies and take corrective actions (e.g., Ployhart & Holtz, 2008). These mean score differences, however, are broadly consistent with other research that has found small, relative to cognitive ability assessments, mean score differences between majority and minority groups in the U.S. on biographical data inventories (Bobko & Roth, 2013; Bobko, Roth, & Potosky, 1999).

Collectively, these results suggest that biographical data inventories may have global utility in the context of high-stakes personnel selection and organizations should consider including these instruments, along with other commonly used and validated predictors, when screening job applicants to improve the strategic value of their selection systems. This research is important because it begins to address calls in the personnel selection literature to shift research from a Western-centric to a multicultural view (Ryan & Ployhart, 2014). In addition, it adds to the limited available literature (e.g., Dalessio et al., 1996; Laurent, 1970) on the cross-national generalizability of biographical data by expanding this research question beyond Europe—to a broader and more diverse sample of countries in Confucian Asia, Southern Asia, and Latin America—and helps clarify conflicting findings in the current literature, which had previously found that biographical data inventories developed in one country (e.g., the United States) may only generalize to culturally similar countries and not to countries that are culturally dissimilar from the country in which it was developed (Hinrichs et al., 1976).

These results are practically important because they suggest that multinational organizations interested in deploying standardized selection systems across geographical boundaries may want to consider including a biographical data inventory to enhance the overall validity of their selection processes. It is worth noting that a single integrated global selection

system is only viable to the extent that there is a common job across geographies. If an organization is able to identify a common job and overcome the challenges related to designing and deploying a global personnel selection system (e.g., candidate reactions, logistical constraints, legal environment, measurement considerations), this approach has the potential to reduce organizations' costs related to developing, implementing, and maintaining selection processes while enabling the organization to efficiently manage human capital (e.g., workforce planning, promotions, identifying high potential employees) by ensuring that all new hires have the necessary KSAOs to succeed on the job. In the terminology of multinational personnel selection strategies, a standardized selection system is aligned with either a global or transnational strategy in that both of these approaches use the same selection tools (transnational) or tools and process (global) across all geographies to help an organization achieve its strategic objectives.

At a more detailed level, these results suggest that a biographical data inventory developed and scored in the U.S. can be successfully generalized to other countries without the added expense of re-validating or creating unique scoring keys. This may allow similar multinational organizations to increase the scale and efficiency of their selection systems by being able to use the same instrument globally, via validity generalization, after translating it into the local language. Second, when an organization uses a rigorously developed competency model as the foundation for its selection instruments, results suggest that a biographical data inventory can successfully assess these KSAOs in both the home country and other geographical regions. Therefore, when combined with other HR policies, practices, and systems, selecting employees on these KSAOs may allow an organization to turn its workforce into a source of competitive advantage.

The current research also revealed a number of practical recommendations for companies interested in similar endeavors. Firstly, it is important to develop three to four times the number of biographical items than desired for the final assessment, as the empirical keying strategy reduced the 206 (108)-item measure to 45 (35) final items for Study 1 (Study 2). Although there is an increased expense associated with creating and piloting additional items, the benefits of scaling a single assessment and creating one scoring algorithm outweighs this cost. Secondly, a large sample size across multiple countries is required to have adequate statistical power to test cross-national comparisons. This is only made possible through the support of upper leadership within an organization; therefore, I urge practitioners to seek buy-in early by highlighting the advantages of this method and its ability to contribute to sustained competitive advantage for the organization. In summary, findings support the cross-national generalizability of biographical data inventories, empirically keyed in the United States, to a broad and diverse sample of country clusters. Given these findings, I echo the sentiments of Ryan and Ployhart (2014) and call other employee selection researchers to action to expand our knowledge of the design and performance of selection instruments outside of the U.S. and Europe.

Limitations

No study is without limitations. Therefore, it is important to acknowledge them in an effort to improve future research on this topic and related areas. First, it was not possible to test for measurement invariance for the biographical data inventory because it was multidimensional, it used a variety of different response formats (e.g., Likert-type, dichotomous), and many countries did not have sufficient sample sizes. The first two characteristics are common to many biographical data inventories and make it difficult to test for measurement invariance because the items are likely to cross-load on multiple factors and the data is likely to violate the assumption

of multivariate normality. The last issue, related to sample size, made it challenging to examine the measurement invariance of both the biographical data inventory and also the job performance measure across country clusters. Future researchers interested in evaluating the measurement invariance of biographical data inventories may consider taking a rational development approach, with consistent response formats, to avoid these data limitations and ensuring enough data is collected during the validation study to evaluate measurement invariance. It is worth noting that differential item functioning (DIF) analyses are “uncommon in the employment domain” (SIOP, 2003, p. 34) because such analyses have rarely found large and replicable DIF effects. Given these findings, SIOP (2003) noted that DIF analyses “should be viewed with caution” and are “not likely to become a routine or expected part of the test validation in employment settings” (p. 34). Based on these considerations, in the context of the current study, there was a greater emphasis on ensuring conceptual equivalence, or the extent to which items and constructs have similar meanings across countries (Lievens et al., 2015). Therefore, it is important for researchers to allocate enough time to comprehensively review and revise the pool of items before beginning data collection.

A second cluster of limitations was related to the sample and design in both studies. All participants were incumbent employees, who worked for a single organization in one industry. Based on this, it is unclear how these findings will generalize to other industries or organizations. The fundamental development approach used in the current studies, however, provides guidance on how other multinational organizations can create biographical data inventories that are likely to exhibit cross-national generalizability. Relatedly, the cross-validation that was conducted on the holdout sample of U.S. participants in both studies was actually a subset of the same U.S. sample rather than an entirely different or independent sample. This dependency, along with

using a concurrent validation design of incumbent employees, likely results in an optimistic estimate of the operational validity of each biographical data inventory (Bliesner, 1996). Furthermore, this research only looked at biographical data inventories in isolation. From a real-world perspective, it is important to understand the cross-national generalizability of a biographical data inventory in the broader context of the end-to-end personnel selection system. That is, does the biographical data consistently provide unique (i.e., incremental validity) information, over other assessment tools such as cognitive ability and personality, across countries? Unfortunately, these questions were outside the scope for the current research and data was unavailable to evaluate them. Given these limitations, it is important for organizations to conduct additional research using predictive validation study designs with applicant samples that incorporate all other selection tools that are part of the hiring process to provide a more realistic appraisal of a selection tool's utility in a global context.

Last, cross-cultural and cross-national research is frequently a challenge due to practical constraints. In Study 1, the biographical data inventory was not translated into the native language for all participants. It is worth noting that all managerial participants had previously demonstrated English Proficiency. In Study 2, however, all forms were translated into the native language of all participants. Looking across the results from both studies, this difference/limitation likely had a minor impact on the overall conclusions, as the results were consist. A second cross-cultural issue with the current research was the use of country as a proxy for culture. While this is convenient and makes sense, given that organizations operate along geographical lines, it is problematic from a research perspective because country and culture are not equivalent. It is possible that there may be multiple subcultures within a single country or political differences within one country (Peterson & Smith, 1997). This limitation would have

been especially problematic if the hypothesis of validity generalization was not supported because it would have been difficult to explain any observed differences between countries. In the context of the current study, however, this limitation was less problematic because the validity generalization hypothesis was supported. Ideally, future researchers should collect data about the cultural values of all participants to verify there is actually a shared set of beliefs within a country.

Future Directions

Researchers interested in the cross-national or cross-cultural use of biographical data inventories should consider looking at alternative biographical data development and scoring methods to understand what conditions inventories are more or less likely to generalize to another country or culture. For example, are rational scoring keys more or less likely to generalize? How does using a work-oriented (vs. a worker-oriented) biographical data development approach influence the cross-national generalizability of biographical data? Does the use of qualitative methods to generate biographical data items influence their global utility? Are some scoring approaches more or less likely to result in cross-nationally generalizable assessments?

Second, additional work is needed to examine the cross-national generalizability of biographical data instruments in the context of end-to-end selection systems. That is, are common U.S. findings related to the incremental validity of biographical data inventories over interviews (Dalessio & Silverhart, 1994), cognitive ability tests (Mael & Hirsch, 1993), and Big Five assessments (Mount, Witt, & Barrick, 2000; Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004) generalizable to other countries or cultures? This is especially important because multinational organizations rarely use a single assessment tool. Therefore, it is important to

ensure that these U.S. results regarding the incremental validity of biographical data inventories and other assessment tools generalize across countries.

Last, in the context of Mael's (1991) taxonomy of biographical data item characteristics, future work is needed to understand what features of items make them more or less likely to generalize to other countries or clusters. How might cultural values influence these results? Is Mael's (1991) recommendation for biographical data items to focus on historical, external, objective, first-hand, discrete, and verifiable behavioral situations equally appropriate in other countries? Researchers interested in evaluating the measurement invariance of biographical data inventories may consider using Mael's (1991) framework to understand why some items may exhibit differential item functioning. Recent analytical advances (e.g., exploratory structural equation modeling) provide a compelling framework to investigate these more focused, item-level questions regarding the cross-national generalizability of biographical data inventories (Morin, Arens, & Marsh, 2016). Further, how do candidates' reactions vary across countries on biographical data items that differ in the extent to which an item favors one or more of these item characteristics? These are important open questions that require additional work to advance the science and practice of multinational personnel selection.

Conclusion

In an increasingly interconnected economy, the labor market has shifted from a domestic to an international perspective. In this environment, multinational organizations demand cost-effective personnel selection tools to help them identify top talent from different countries, geographical regions, and cultures. The purpose of the current research was to evaluate the global utility of biographical data inventories, a standardized self-report selection method that asks job applicants questions about prior behaviors and experiences. Results from two studies

involving participants from 7 country clusters, across four continents, and two different occupational groups, managers and manufacturing technicians, provided evidence to support the hypothesis that biographical data inventories, empirically keyed in the United States, generalize to all country clusters examined. This research is important because it helps to clarify the current literature on the cross-national generalizability of biographical data inventories and provides multinational organizations a potential way to enhance the effectiveness of their selection systems. Future research should continue to take a multicultural view of personnel selection to ensure that Industrial-Organizational Psychology research helps to inform HR practices in global organizations.

References

- 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.
- Aguinis, H., Petersen, S. A., & Pierce, C. A. (1999). Appraisal of the homogeneity of error variance assumption and alternatives to multiple regression for estimating moderating effects of categorical variables. *Organizational Research Methods, 2*(4), 315-339.
- Aguinis, H., & Pierce, C. A. (1998). Heterogeneity of error variance and the assessment of moderating effects of categorical variables: A conceptual review. *Organizational Research Methods, 1*(3), 296-314.
- Aguinis, H., Pierce, C. A., & Stone-Romero, E. F. (1994). Estimating the power to detect dichotomous moderators with moderated multiple regression. *Educational and Psychological Measurement, 54*, 690-692.
- Akamai. (2014). The state of the Internet report. <http://www.akamai.com/stateoftheinternet/>.
- Alexander, R. A., & Govern, D. M. (1994). A new and simpler approximation for ANOVA under variance heterogeneity. *Journal of Educational Statistics, 19*, 91-101.
- Anderson, N., Salgado, J. F., & Hülsheger, U. R. (2010). Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment, 18*(3), 291–304.
- Arthur, W. J., & 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*(2), 435–442.

- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1–26.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment, 9*(1–2), 9–30.
- Bartram, D. (2004). Assessment in organizations. *Applied Psychology: An International Review, 53*(2), 237–259.
- Beatty, A. S. (2013). *A critical review of empirical and rational strategies for item selection and keying for biographical data inventories* (Unpublished doctoral dissertation). University of Minnesota, Minneapolis, MN.
- Bergman, M. E., Drasgow, F., Donovan, M. A., Henning, J. B., & Juraska, S. E. (2006). Scoring situational judgment tests: Once you get the data, your troubles begin. *International Journal of Selection and Assessment, 14*(3), 223–235.
- Binning, J. F., & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology, 74*(3), 478–494.
- Bliesener, T. (1996). Methodological moderators in validating biographical data in personnel selection. *Journal of Occupational and Organizational Psychology, 69*, 107–120.
- Bobko, P., & Roth, P. L. (2013). Reviewing, categorizing, and analyzing the literature on black-white mean differences for predictors of job performance: Verifying some perceptions and updating/correcting others. *Personnel Psychology, 66*, 91–126.
- Bobko, P., Roth, P. L., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology, 52*, 561–589.

- Brannick, M. T., Levine, E. L., & Morgeson, F. P. (2007). *Job and work analysis: Methods, research, and applications for human resource management* (2nd ed.). Thousand Oaks, CA: Sage.
- Breaugh, J. A. (2009). The use of biodata for employee selection: Past research and future directions. *Human Resource Management Review, 19*, 219–231.
- Briscoe, D. R. (1997). Assessment centers: Cross-cultural and cross-national issues. *Journal of Social Behavior and Personality, 12*, 261–270.
- Caligiuri, P., & Paul, K. B. (2010). Selection in multinational organizations. In J. L. Farr & N. T. Tippins (Eds.), *The Handbook of Employee Selection* (pp. 781–799). New York, NY: Routledge.
- Carey, T., Herst, D., & Chan, W. (2010). Global selection: Selection in international contexts. In K. Lundby & J. Jolton (Ed.), *Going Global: Practical Applications and Recommendations for HR and OD Professionals in the Global Workplace*. (pp. 143–74). San Francisco, CA: Jossey-Bass.
- Carlson, K. D., Scullen, S. E., Schmidt, F. L., Rothstein, H., & Erwin, F. (1999). Generalizable biographical data validity can be achieved without multi-organizational development and keying. *Personnel Psychology, 52*(3), 731–755.
- Cascio, W. F., & Aguinis, H. (2011). *Applied psychology in human resource management* (7th Ed.). Upper Saddle River, NJ: Prentice Hall.
- Chan, D. (2008). Methodological issues in international human resource management research. In M. M. Harris (Ed.), *Handbook of Research in International Human Resources Management* (pp. 53–76). New York, NY: Lawrence Erlbaum Associates.

- Chaney, F. B., & Owens, W. A. (1964). Life history antecedents of sales, research, and general engineering interest. *Journal of Applied Psychology, 48*(2), 101-105.
- Chen, C., Lee, S., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science, 6*(3), 170–175.
- Chen, P. Y., & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures*. Thousand Oaks, CA: Sage Publications, Inc.
- Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and White students in integrated colleges. *Journal of Educational Measurement, 5*(2), 115–124.
- Combs, J., Liu, Y., Hall, A., & Ketchen, D. (2006). How much do high-performance work practices matter? A meta-analysis of their effects on organizational performance. *Personnel Psychology, 59*, 501–528.
- Crook, T. R., Todd, S. Y., Combs, J. G., Woehr, D. J., & Ketchen, D. J. (2011). Does human capital matter? A meta-analysis of the relationship between human capital and firm performance. *Journal of Applied Psychology, 96*(3), 443–456.
- Cucina, J. M., Caputo, P. M., Thibodeaux, H. F., & Maclane, C. N. (2012). Unlocking the key to biodata scoring: A comparison of empirical, rational, and hybrid approaches at different sample sizes. *Personnel Psychology, 62*, 385–428.
- Cucina, J. M., & Walmsley, P. T. (2015, April). *Empirically-keyed biodata follows the scientific method: 'Atheoretical' need not apply*. Presented at the 30th annual conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

- Dalessio, A. T., Crosby, M. M., & McManus, M. A. (1996). Stability of biodata keys and dimensions across English-speaking countries: A test of the cross-situational hypothesis. *Journal of Business and Psychology, 10*(3), 289–296.
- Dalessio, A. T., & Silverhart, T. D. (1994). Combining biodata test and interview information: Predicting decisions and performance criteria. *Personnel Psychology, 47*, 303–315.
- Dean M. A., Russell C. J., Farmer W. (2002). Noncognitive predictors of air traffic controller performance. In H. Eibfeldt, M. Heil, D. Broach, (Eds.), *Staffing the air traffic management (ATM) system: The selection of air traffic controllers* (pp. 59–72). Brookfield, VT: Ashgate.
- Dean, M. A., Russell, C. J., & Muchinsky, P. M. (1999). Life experiences and performance prediction: Toward a theory of biodata. In G. Ferris (Ed.), *Research in Personnel/Human Resources Management* (pp. 245-281). Greenwich, CT: JAI Press.
- DeShon, R. P., & Alexander, R. A. (1996). Alternative procedures for testing regression slope homogeneity when group error variances are unequal. *Psychological Methods, 1*, 261-277.
- Erez, M. (2011). Cross-cultural and global issues in organizational psychology. In S. Zedeck (Ed.), *APA Handbook of Industrial and Organizational Psychology, Vol. 3: Maintaining, Expanding, and Contracting the Organization* (pp. 807–854). Washington, DC, US: American Psychological Association.
- Erwin, F. W., & Herring, J. W. (1977). *The feasibility of the use of autobiographical information as a predictor of early Army attrition*. Alexandria, VA: U.S. Army Institute for Behavioral and Social Sciences.

- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. -G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160.
- Field, A. (2005). *Discovering statistics using SPSS*. Thousand Oaks, CA: Sage.
- Friedman, T. L. (2007). *The world is flat 3.0: A brief history of the twenty-first century*. New York, NY: Picador.
- Furnham, A. (2008). HR professionals' beliefs about, and knowledge of, assessment techniques and psychomotor tests. *International Journal of Selection and Assessment*, *16*(3), 300–305.
- Hinrichs, J. R., Haanperä, S., & Sonkin, L. (1976). Validity of a biographical information blank across national boundaries. *Personnel Psychology*, *29*, 417–421.
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Application reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, *57*, 639–683.
- Herriot, P., & Anderson, N. (1997). Selecting for change: How will personnel selection psychology survive? In N. Anderson & P. Herriot (Eds.), *International handbook of selection and assessment* (pp.1–34). London, UK: Wiley.
- Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Newbury Park, CA: Sage.
- Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations*. Thousand Oaks, CA: Sage.
- Hough, L. M. (2010). Assessment of background and life experience: The past as prologue. In J. C. Scott, & D. H. Reynolds (Eds.), *Handbook of workplace assessment: Evidence-based*

- practices for selecting and developing organizational talent* (pp. 109-139). San Francisco, CA: Jossey-Bass.
- Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, 75(5), 581–595.
- House, R. J., Hanges, P. W., Javidan, M., Dorfman, P., & Gupta, V. (Eds.). (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Thousand Oaks, CA: Sage.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Hui, C. H., & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*, 20(3), 296–309.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96(1), 72–98.
- International Test Commission (ITC). (2005). *ITC guidelines for translating and adapting tests*. Retrieved from https://www.intestcom.org/files/guideline_test_adaptation.pdf.
- James, G. S. (1951). The comparison of several groups of observations when the ratios of population variances are unknown. *Biometrika*, 38, 324-329.
- Kluger, A. N., Reilly, R. R., & Russell, C. J. (1991). Faking biodata tests: Are option-keyed instruments more resistant? *Journal of Applied Psychology*, 76(6), 889–896.
- Kristof, A. L. (1996). Person–organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology*, 49, 1– 49.

- Laurent, H. (1970). Cross-cultural cross-validation of empirically validated tests. *Journal of Applied Psychology, 54*(5), 417–423.
- Lautenschlager, G. J., & Mendoza, J. L. (1986). A step-down hierarchical multiple regression analysis for examining hypotheses about test bias in prediction. *Applied Psychological Measurement, 10*(2), 133–139.
- Lievens, F. (2008). Research on selection in an international context: Current status and future directions. In M.M. Harris (Ed.), *Handbook of Research in International Human Resource Management* (pp. 107–123). Mahwah, NJ: Lawrence Erlbaum.
- Lievens, F., Corstjens, J., Sorrel, M. A., Abad, F. J., Olea, J., & Ponsoda, V. (2015). The cross-cultural transportability of situational judgment tests: How does a US-based integrity situational judgment test fare in Spain? *International Journal of Selection and Assessment, 23*(4), 361–372.
- Lievens, F., Harris, M. M., Van Keer, E., & Bisqueret, C. (2003). Predicting cross-cultural training performance: The validity of personality, cognitive ability, and dimensions measured by an assessment center and a behavior description interview. *Journal of Applied Psychology, 88*(3), 476–489.
- Lievens, F., Sanchez, J. I., & De Corte, W. (2004). Easing the inferential leap in competency modeling: The effects of task-related information and subject matter expertise. *Personnel Psychology, 57*(4), 881–904.
- Mael, F. A. (1991). A conceptual rationale for the domain and attributes of biodata items. *Personnel Psychology, 44*, 763–792.
- Mael, F. A., & Hirsch, A. C. (1993). Rainforest empiricism and quasi-rationality: Two approaches to objective biodata. *Personnel Psychology, 46*, 719–738.

- Millsap, R. E., & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement, 17*(4), 297–334.
- Mitchell, T. W., & Klimoski, R. J. (1982). Is it rational to be empirical? A test of methods for scoring biographical data. *Journal of Applied Psychology, 67*(4), 411–418.
- Motowidlo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. *Human Performance, 10*(2), 71–83.
- Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*(6), 640–647.
- Morin, A. J. S., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(1), 116–139.
- Mount, M. K., Witt, L. A., & Barrick, M. R. (2000). Incremental validity of empirically keyed biodata scales over GMA and the Five Factor personality constructs. *Personnel Psychology, 53*(2), 299–323.
- Mumford, M. D., Barrett, J. D., & Hester, K. S. (2012). Background data: Use of experiential knowledge in personnel selection. In N. Schmitt (Ed.), *The Oxford Handbook of Personnel Assessment and Selection* (pp. 353–382). New York, NY: Oxford University Press.
- Mumford, M. D., Costanza, D. P., Connelly, M. S., & Johnson, J. F. (1996). Item generation procedures and background data scales: Implications for construct and criterion-related validity. *Personnel Psychology, 49*(2), 361–398.

- Mumford, M. D., & Owens, W. A. (1987). Methodology review: Principles, procedures, and findings in the application of background data measures. *Applied Psychological Measurement, 11*(1), 1–31.
- Mumford, M. D., Stokes, G. S., & Owens, W. A. (1990). *Patterns of life history: The ecology of human individuality*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Mumford, M. D., Uhlman, C. E., & Kilcullen, R. N. (1992). The structure of life history: Implications for the construct validity of background data scales. *Human Performance, 5*(1-2), 109–137.
- Nickels, B. J. (1994). The nature of biodata. In G. S. Stokes, M. D., Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 1–16). Palo Alto, CA: Consulting Psychologists Press.
- Oh, I.-S. (2009). *The Five-Factor Model of personality and job performance in East Asia: A cross-cultural validity generalization study* (Unpublished doctoral dissertation). University of Iowa, Iowa City, IA.
- Oh, I.-S., Guay, R. P., Kim, K., Harold, C. M., Lee, J.-H., Heo, C.-G., & Shin, K.-H. (2014). Fit happens globally: A meta-analytic comparison of the relationships of person-environment fit dimensions with work attitudes and performance across East Asia, Europe, and North America. *Personnel Psychology, 67*, 99–152.
- Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science, 7*(6), 657-660.

- Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology, 89*(2), 187–207.
- Peterson, M. F., & Smith, P. B. (1997). Does national culture or ambient temperature explain cross-national differences in role stress? No sweat! *Academy of Management Journal, 40*(4), 930–946.
- Ployhart, R. E., & Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology, 61*, 153-172.
- Ployhart, R. E., & Moliterno, T. P. (2011). Emergence of the human capital resource: A multilevel model. *Academy of Management Review, 36*(1), 127–150.
- Ployhart, R. E., & Weekley, J. A. (2010). Strategy, selection, and sustained competitive advantage. In J. L. Farr & N. T. Tippins (Eds.), *The Handbook of Employee Selection* (pp. 195–212). New York, NY: Routledge.
- Ployhart, R. E., & Weekley, J. A. (2015). Recruitment and selection in global organizations: A resource-based framework. In D. G. Collings, G. T. Wood, & P. M. Caligiuri (Eds.), *The Routledge Companion to International Human Resource Management* (pp. 155–171). New York, NY: Routledge.
- Popper, K. R. (1959). *The logic of scientific discovery*. London, United Kingdom: Hutchinson.
- Ryan, A. M., McFarland, L., Baron, H., & Page, R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology, 52*, 359–391.

- Ryan, A. M., & Ployhart, R. E. (2014). A century of selection. *Annual Review of Psychology*, *65*, 693–717.
- Ryan, A. M., & Tippins, N. T. (2004). Attracting and selecting: What psychological research tells us. *Human Resource Management*, *43*(4), 305–318.
- Ryan, A. M., & Tippins, N. (2009). *Designing and implementing global selection systems*. West Sussex, U.K: Wiley-Blackwell.
- Ryan, A. M., Wiechmann, D., & Hemingway, M. (2003). Designing and implementing global staffing systems: Part II-Best practices. *Human Resource Management*, *42*(1), 85–94.
- Sackett, P. R., Shen, W., Myors, B., Lievens, F., Schollaert, E., Van Hoye, G., et al. (2010). Perspectives from twenty-two countries on the legal environment for selection. In J. L. Farr & N. T. Tippins (Eds.), *The Handbook of Employee Selection* (pp. 651–676). New York, NY: Routledge.
- Salas, E., Tannenbaum, S. I., Kraiger, K., & Smith-Jentsch, K. A. (2012). The science of training and development in organizations: What matters in practice. *Psychological Science in the Public Interest*, *13*(2), 74–101.
- Salgado, J. F. (1997). The Five Factor Model of personality and job performance in the European Community. *Journal of Applied Psychology*, *82*(1), 30–43.
- Salgado, J. F. (1998a). Big Five personality dimensions and job performance in army and civil occupations: A European perspective. *Human Performance*, *11*(2), 271–288.
- Salgado, J. F. (1998b). Sample size in validity studies of personnel selection. *Journal of Occupational and Organizational Psychology*, *71*, 161-164.
- Salgado, J., & Anderson, N. (2002). Cognitive and GMA testing in the European Community: Issues and evidence. *Human Performance*, *15*(1/2), 75–96.

- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C. & de Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology*, *56*, 573–605.
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F., & Rolland, J. P. (2003). A meta-analytic study of general mental ability validity for different occupations in the European Community. *Journal of Applied Psychology*, *88*(6), 1069–1081.
- Salgado, J. F., Viswesvaran, C., & Ones, D. S. (2001). Predictors used for personnel selection: An overview of constructs, methods and techniques. In N. Anderson, D.S. Ones, H.K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of Industrial, Work and Organizational Psychology* (Vol. 1, pp. 165–199). London, England: Sage.
- Sanchez, J. I., & Levine, E. L. (2009). What is (or should be) the difference between competency modeling and traditional job analysis? *Human Resource Management Review*, *19*, 53–63.
- Schaffer, B. S., & Riordan, C. M. (2003). A review of cross-cultural methodologies for organizational research: A best-practices approach. *Organizational Research Methods*, *6*(2), 169–215.
- Schippmann, J. S., Ash, R. A., Battista, M., Carr, L., Eyde, L. D., Hesketh, B., . . . Sanchez, J. I. (2000). The practice of competency modeling. *Personnel Psychology*, *53*(3), 703–740.
- Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, *62*(5), 529–540.
- 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*(2), 262–274.

- Schmitt, N., Cortina, J. M., Ingerick, M. J., & Wiechmann, D. (2003). Personnel selection and employee performance. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of Psychology: Vol. 12 Industrial and Organizational Psychology* (pp.77–105). Hoboken, NJ: Wiley.
- Schmitt, N., & Golubovich, J. (2013). Biographical information. In K. F. Geisinger (Ed.), *APA Handbook of Testing and Assessment in Psychology: Vol. 1. Test Theory and Testing and Assessment in Industrial and Organizational Psychology* (pp. 437–455). Washington, DC: American Psychological Association.
- Schwartz, S. H. (1994). Beyond individualism/collectivism: New cultural dimensions of values. In U. Kim, H. C. Triandis, C. Kâğıtçıbaşı, S. Choi, & G. Yoon (Eds.), *Individualism and collectivism: Theory, method, and applications* (Cross-cultural research and methodology series, Vol. 18, pp. 85–119). Thousand Oaks, CA: Sage.
- Shen, W., Kiger, T. B., Davies, S. E., Rasch, R. L., Simon, K. M., & Ones, D. S. (2011). Samples in applied psychology: Over a decade of research in review. *Journal of Applied Psychology, 96*(5), 1055–1064.
- Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures*. Bowling Green, OH: Society for Industrial and Organizational Psychology.
- Spector, P. E., Liu, C., & Sanchez, J. I. (2015). Methodological and substantive issues in conducting multinational and cross-cultural research. *Annual Review of Organizational Psychology and Organizational Behavior, 2*(1), 101–131.

- Taras, V., Kirkman, B. L., & Steel, P. (2010). Examining the impact of culture's consequences: A three-decade, multilevel, meta-analytic review of Hofstede's cultural value dimensions. *Journal of Applied Psychology, 95*(3), 405–439.
- Terpstra, D. E., Mohamed, A. A., Kethley, R. B. (1999). An analysis of federal court cases involving nine selection devices. *International Journal of Selection and Assessment, 7*(1), 26–34.
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44*(4), 703–742.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70.
- Vinchur, A. J., Schippmann, J. S., & Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology, 83*(4), 586–597.
- Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology, 90*(1), 108–131.
- Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology, 52*(5), 372–376.
- Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology, 67*, 189–205.
- Wright, P. M., Dunford, B. B., & Snell, S. A. (2001). Human resources and the resource based view of the firm. *Journal of Management, 27*, 701–721.

Wright, P. M., McMahan, G. C., & McWilliams, A. (1994). Human resources and sustained competitive advantage: A resource-based perspective. *International Journal of Human Resource Management*. 5(2), 301–326.