



More alike than different? A comparison of variance explained by cross-cultural models

James G. Field¹,
Frank A. Bosco²,
David Kraichy³,
Krista L. Uggerslev⁴ and
Mingang K. Geiger¹

¹Department of Management, John Chambers College of Business and Economics, West Virginia University, Morgantown, WV 26506, USA;

²Department of Management and Entrepreneurship, School of Business, Virginia Commonwealth University, Richmond, VA 23284, USA; ³Department of Human Resources and Organizational Behaviour, Edwards School of Business, University of Saskatchewan, Saskatoon, SK S7N 5A7, Canada; ⁴Applied Research Chair in Leadership and Talent, Office of Research and Innovation, Northern Alberta Institute of Technology, Edmonton, AB T5G 2R1, Canada

Correspondence:

JG Field, Department of Management, John Chambers College of Business and Economics, West Virginia University, Morgantown, WV 26506, USA
e-mail: james.field2@mail.wvu.edu

Abstract

Relatively little is known about the extent to which culture moderates findings in applied psychology research. To address this gap, we leverage the metaBUS database of over 1,000,000 published findings to examine the extent to which six popular cross-cultural models explain variance in findings across 136 bivariate relationships and 56 individual cultural dimensions. We compare moderating effects attributable to Hofstede's dimensions, GLOBE's practices, GLOBE's values, Schwartz's Value Survey, Ronen and Shenkar's cultural clusters, and the United Nations' M49 standard. Results from 25,296 multilevel meta-analyses indicate that, after accounting for statistical artifacts, cross-cultural models explain approximately 5–7% of the variance in findings. The variance explained did not vary substantially across models. A similar set of analyses on observed effect sizes reveal differences of $|r| = .05$ – $.07$ attributable to culture. Variance among the 136 bivariate relationships was explained primarily by sampling error, indicating that cross-cultural moderation assessments require atypically large sample sizes. Our results provide important information for understanding the overall level of explanatory power attributable to cross-cultural models, their relative performance, and their sensitivity to variance in the topic of study. In addition, our findings may be used to inform power analyses for future research. We discuss implications for research and practice.

Journal of International Business Studies (2021) **52**, 1797–1817.

<https://doi.org/10.1057/s41267-021-00428-z>

Keywords: meta-analysis; big data; open science; cross-cultural research/measurement issues

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1057/s41267-021-00428-z>.

Received: 15 June 2020

Revised: 8 March 2021

Accepted: 11 March 2021

Online publication date: 23 April 2021

INTRODUCTION

Cross-cultural and international business research has helped to shape theory in organizational science and to inform evidence-based practice. In terms of theory, culture is often considered to be a boundary condition for phenomena observed in organizational research (e.g., Zhong et al., 2016) and, thus, can describe generalizability limits of a conceptual framework (e.g., Busse et al., 2017). Regarding practice, research has demonstrated that interventions targeting dimensions of job embeddedness (e.g., person–job fit) are valuable for employee retention in different cultures (Ramesh & Gelfand, 2010). The importance of understanding cultural differences is also illustrated by the presence of large-scale reviews of

proposed cultural dimensions (e.g., individualism; Kirkman et al., 2006). In fact, follow-ups of the reviews themselves have appeared (Kirkman et al., 2017). These contributions, as well as a growing need to better understand organizational science theories through a cross-cultural lens, add credence to the claim that “cross-cultural research is needed more than ever” (Gelfand et al., 2017, p. 14).

Culture is a multifaceted concept that concerns values, beliefs, attitudes, and behavior (Taras, Kirkman, & Steel, 2010a). A primary objective of research on culture is to understand if and how organizational behavior differs around the globe (Tsui et al., 2007). To this end, cross-cultural researchers may study the relationship between individualism and turnover intention (i.e., culture-as-main effect at the individual level of analysis) or the relationship between stress and turnover intention as moderated by individualism (i.e., culture-as-moderator at the individual level of analysis). Indeed, findings in this scientific space are most frequently expressed at the individual- (80%) or country-level (17%) (Taras et al., 2010a). Cross-cultural research at the individual level primarily operationalizes culture as a main effect (74%), as opposed to a culture as a moderator effect (26%) (Kirkman et al., 2006). Drawing on the results of almost 600 studies, Taras et al. (2010a) concluded that culture-as-main effects are typically small. Yet, to the best of our knowledge, an analogue of Taras et al.’s study for culture-as-moderator effects has not been published. Indeed, this is surprising given that effect sizes are “what science is all about” (Cohen, 1988: p. 532), and an awareness of how effect sizes vary around the globe will help to contextualize applied psychology theory and evidence-based practice recommendations.

Although interest in cross-cultural research is increasing (Gelfand et al., 2017; Tsui et al., 2007), there remain differences between scholars regarding the measurement and conceptualization of culture (Tung & Verbeke, 2010). For example, one measurement criticism concerns the use of self-report questionnaires (e.g., response-style effects; Cheung & Rensvold, 2000). Similarly, as Matsumoto (2006: 33) asked, “do verbal descriptions about culture reflect anything other than verbal descriptions of culture?” Nonetheless, theories have proliferated (Gelfand et al., 2017) that differ in terms of their explanatory mechanisms (e.g., values, practices, geography). Given this proliferation, a large-scale analysis of the perspectives’ relative performance is warranted.

In the current study, our goal is to advance international business and cross-cultural theory by competitively evaluating four meta-theoretical perspectives on culture. Specifically, inspired by the “strong inference” approach (Platt, 1964), we use a database of more than 1,000,000 published findings to examine the explanatory power of six aggregated (e.g., Hofstede’s six dimensions) and 56 specific (e.g., Hofstede’s individualism) cultural dimensions across 136 substantive bivariate relationships commonly investigated in applied psychology (e.g., job satisfaction–turnover intention). The nature of our study is explanatory, which is why we do not advance specific directional hypotheses regarding the predictive validity of the perspectives under investigation in the current study. Indeed, it is not our goal to definitively establish the best way to think about culture, but to simply examine how much culture impacts bivariate relationships, and to what extent the impact varies across perspectives. To this end, we approach three research questions.

First, we ask: *How well do different cross-cultural models explain variance in published research findings?* To answer this research question, we located findings to be used as input to 136 meta-analyses and examined the extent to which six aggregated models of culture (e.g., Hofstede 1980 vs. Schwartz 1992, 2006) explain observed variance. In these analyses, we control for a variety of statistical artifacts (e.g., unreliability, temporal trends, cultural dynamics; Kashima et al., 2019), and also nuance the analyses by each model’s specific cultural dimensions (e.g., Hofstede’s 1980 individualism vs. Schwartz’s 2006 harmony). Answers to this question serve as a barometer for scientific progress with which to compare model performance.

Our second research question asks: *What is the magnitude of observed effect size differences due to culture?* Although seemingly similar to our first research question, this question is tailored to inform power analysis. That is, we ask how large an observed effect size difference one might expect if, for example, studying the relationship between job satisfaction and performance in Anglo versus non-Anglo cultures. Obtaining this information is important for researchers planning future cross-cultural studies. Like our first research question, we nuance these analyses by specific dimension.

Finally, for our third research question we ask: *To what extent are 136 bivariate relationships differentially moderated by culture?* Put differently, we aim to ascertain the extent to which certain bivariate

relationships present with stronger or weaker cross-cultural effects. In addressing this question, we also consider the possibility that the variance in the 136 moderating effects is due not to the bivariate relationship but instead explained by sampling error.

The remainder of the present study is organized as follows. We begin by describing the models of culture and corresponding cultural dimensions that are under investigation. Second, we report results from a series of multilevel meta-analyses that collectively assess the extent to which culture moderates substantive bivariate relationships. Third, in our Discussion, we describe the implications of our findings for research and practice. In addition, we introduce web-based software called CASST (Cross-Area Sample Size Tool; see <https://casst.shinyapps.io/gen1/>) that allows readers to interact with our results, and can be used by future cross-cultural researchers to aid with a priori sample size estimation. Finally, we provide limitations of our study and outline how future research may address them.

META-THEORETICAL PERSPECTIVES ON CULTURE

In this section, we describe four meta-theoretical perspectives that are commonly used to inform and test hypotheses on cross-cultural effects in applied psychology research: cultural values dimensions, cultural practices dimensions, configurations, and geographic regions. Due to space constraints, we are unable to provide an exhaustive description of all perspectives in this research area. We broadly describe how each meta-theoretical perspective is used to conceptualize and test cross-cultural effects, and illustrate fundamental differences between each perspective, rather than wading into the operational and conceptual nuances.

Perhaps the most widely adopted meta-theoretical approach for testing cross-cultural differences is the *cultural values* perspective (Taras et al., 2010a). This perspective generally focuses on ostensive values differences between nations and societies as characterized in Hofstede's (1980) seminal study, in later works by Schwartz (1992, 2006), and in the Global Leadership and Organizational Behavior Effectiveness (GLOBE; House et al., 2004). The first iteration of this perspective was guided by a country-level factor analysis using attitudinal survey data collected from IBM and its subsidiaries in 72 countries between 1968 and 1973 (Hofstede, 1980).

Hofstede's results allowed the classification of countries along four cultural dimensions (i.e., individualism–collectivism, power distance, uncertainty avoidance, and masculinity–femininity). Subsequent studies (e.g., Hofstede et al., 2010) resulted in two additional cultural dimensions (i.e., long-term orientation and indulgence–restraint).

Despite its widespread use, scholars have raised several concerns regarding Hofstede's classification approach (e.g., Minkov, 2018). Specifically, scholars have argued that it is too "blunt of an object for picking up on certain cultural nuances" (Shao et al., 2013: 283; see also House et al., 2004). These criticisms led researchers to develop revised models in an attempt to better explain cross-cultural differences. For example, Schwartz's (1992, 2006) theories on cultural value orientations and personal values are often used to develop and test hypotheses on cross-cultural effects at the national and individual level, respectively. Although Schwartz's and Hofstede's typologies of cultural values are non-orthogonal, and both approaches rely on aggregated self-report data to classify national cultures, the former is different from the latter because it specifies its structure of cultural dimensions in terms of the shared *and* opposing assumptions that underlie them. Likewise, House et al.'s (2004) GLOBE project used implicit leadership, implicit motivation, and value-belief theories to develop nine dimensions of societal and organizational culture. Although these popular models of cultural values partially overlap, and each approach relies on aggregated self-report data to classify national cultures, there exist substantial conceptual, operational, and measurement differences that separate them.

It has been suggested that values may predict one type of behavior in one culture and a different type of behavior in another culture (Knafo et al., 2011). Furthermore, it has been argued that values and behaviors are not always isomorphic, meaning that the way things are done does not always match the ideal way of doing things (House et al., 2004). As such, to make better sense of a culture we need to know what values (i.e., "what should be") *and* local behaviors (i.e., "what is") are shared across cultures. This corollary notion is captured by the *cultural practices dimensions* perspective, which stems from psychological and behavioral traditions in which it is argued that cultures should be studied as they are construed by their members (e.g., Segall et al., 1998). The GLOBE project's cultural practices

dimensions capture the behavioral manifestation of culture; the practices dimensions scales were adapted from the corresponding values dimensions scales. Although meta-analytic evidence indicates that GLOBE's cultural values dominions may have greater explanatory power than the corresponding cultural practices dimensions (e.g., Crede et al., 2019), we are not aware of any large-scale study that simultaneously compares the predictive validity of both types of meta-theoretical perspectives (i.e., cultural values dimensions vs. cultural practices).

Although there is a growing body of evidence that suggests that national borders are not the best demarcation to understand cross-cultural effects (e.g., Taras et al., 2016), studies that report how meta-analytic mean effect size estimates vary between individual countries continue to appear in the published literature. One criticism of the *geography* perspective is that cultural differences are unlikely to be detected with sufficient power when substantive bivariate relationships are compared between individual countries. Indeed, recent evidence indicates that inferences regarding the impact of cultural dimensions on substantive relationships should not be formulated when culture scores from fewer than seven countries are analyzed, which adds credence to the claim that the "passport" approach to cross-cultural research should be abandoned (Franke & Richey, 2010). The United Nations' M49 Standard strictly geographically-based categorization approach (United Nations Statistics Division, 1999), which arranges countries into 17 sub-regions (e.g., Northern America), is one country classification system that may help to circumvent several concerns levied against national comparisons.¹ However, given that the M49 and similar geographic clustering approaches are rarely, if ever, used by cross-cultural researchers, their efficacy for explaining variance in cross-cultural effects is not well understood.

Given that culture can be conceived as a latent, hypothetical construct that represents a complex pattern of relationships among a plethora of cultural dimensions (Tsui et al., 2007), a common criticism of the aforementioned cultural values and cultural practices perspectives is that they focus on using predominantly a single cultural dimension (i.e., individualism), or multiple cultural dimensions independently, to explain differences across cultures (e.g., Fisher, 2014). Such shortcomings led scholars to develop the *configuration* perspective (e.g., Beugelsdijk et al., 2017; Ronen & Shenkar,

2013). Under this meta-theoretical perspective, cultures are not arranged according to self-response data from surveys intended to measure social constructions of what is believed to be culture, but rather profiles of relatively objective ecocultural antecedents (e.g., religion, language) and economic corollaries (e.g., economic freedom) are used to organize countries into theoretically relevant shared cultural clusters. One benefit of this approach is that it provides the opportunity to explore and test the viability of new regional-theories, and the cultural relativity of existing organizational theories at an alternate unit of analysis (i.e., supranational vs. country level). Furthermore, the configuration perspective is attractive because cultural clusters are determined based on the extent to which countries are similar along *multiple* relatively objective dimensions. Indeed, this suggests that the configuration perspective should help to explain more variance in cross-cultural effects than perspectives that rely on single value, behavioral, or geographic comparisons. However, whether or not this is the case has not been determined.

Taken together, there is a growing interest in cross-cultural effects (Gelfand et al., 2017), which is reflected in several meta-theoretical perspectives that endeavor to explain how and why substantive relationships might differ across cultures. Yet, evidence accumulated over 40 years of cross-cultural research has not been synthesized to ascertain the extent to which culture matters (e.g., Allen et al., 2015) and whether popular cross-cultural models exhibit a validity ceiling (cf. Cascio & Aguinis, 2008).

PRESENT STUDY

In the present study, we examine the extent to which variance in culture-as-moderator effects are driven by emic (i.e., culture-specific; Gelfand, Raver, & Holcombe Ehrhart, 2002), methodological (e.g., sampling error), or substantive (e.g., bivariate relationship) factors. Furthermore, answers to the our research questions will offer information that can be used to (1) conduct better informed a priori statistical power analyses, which may help to mitigate underpowered studies and reduce Type II errors in cross-cultural research (Maxwell, 2004), (2) better understand and interpret the magnitude of cross-cultural effects (cf. Taras et al., 2010a), and (3) provide cross-cultural researchers with an alternative perspective through which to compare the

explanatory power of popular cross-cultural models.

Our study contributes to the extant literature by responding to a number of recent recommendations for cross-cultural research. First, our study “move[s] beyond studying culture at the national level” (Gelfand et al., 2017: 521) by using a variety of multi-dimensional models to examine the prevalence and magnitude of cross-cultural effects. Second, we answer Teagarden, Von Glinow, and Mellahi’s call to “integrate context” (2018: p. 304) by examining the degree to which culture-as-moderator effects vary across 136 commonly investigated relationships in applied psychology research. Third, we respond to calls to adopt a “big data” approach to measuring the effect of culture (Gelfand et al., 2017) by drawing on a database of over 1,000,000 effect sizes (Bosco et al., 2017) to answer our research questions. Fourth, in response to calls to improve methodological robustness in cross-area research (Aguinis et al., 2017), we introduce an open-access application that allows users to interact with the observed results and make our study materials available (see <http://osf.io/32c4y/>) so that other researchers can replicate and extend our findings.

METHODS

Database

For all analyses, we relied on the metaBUS database (v.2017.10.23) containing 1,038,319 effect sizes (i.e., correlation coefficients) reported in 16,616 articles between 1980 and 2017 (Bosco et al., 2017). Manually coded database fields used in the present study include (1) study sample size, (2) country of data collection, and (3) variable classification according to Bosco et al.’s (2015) hierarchical taxonomy. For a detailed description of the metaBUS database contents, structure, and coding protocol, see Bosco et al. (2017).

We sought to conduct our analyses on a broad array of common bivariate relationships that represent major applied psychology topics. To this end, we identified and refined a list of 17 constructs including common behaviors (e.g., in-role performance, extra-role performance, counterproductive work behavior), attitudes (e.g., job satisfaction, organizational commitment), psychological traits (e.g., conscientiousness, extraversion), and states (e.g., happiness, anger [nested beneath positive and negative states, respectively]), among other

constructs. Although relationships involving some demographic variables (e.g., age, sex) are reported with high frequency, we chose to omit these variables for two reasons. First, such variables are often reported in matrices by convention rather than by dint of research foci or hypotheses. Second, effects involving demographic variables tend to be substantially weaker than those for other sorts of bivariate relationships (Bosco et al., 2015), which could yield downwardly biased estimates of the cultural moderation effects studied here.

As shown in Table 1, combinations of the 17 constructs produce a matrix of 136 bivariate relationships that together summarize 50,294 published findings. Importantly, in contrast to existing large-scale summaries of cross-cultural research, our goal was to avoid bivariate relationships with cultural constructs (e.g., individualism). Indeed, bivariate relationships involving cultural constructs would represent “main” effects (e.g., the relationship between individualism and job satisfaction), and our purpose was to investigate moderating effects (e.g., the relationship between job satisfaction and in-role performance as moderated by individualism).

In the metaBUS database, sample size is operationalized as the minimum of each variable pair’s sample sizes. For example, imagine a matrix wherein the sample size for job satisfaction is 200 and the sample size for autonomy is 180 due to missing data. In this case, the coded sample size for the variable pair is 180.

Cross-cultural Models

We sought to compare moderating effect sizes attributable to a variety of cross-cultural models. To this end, we selected several cross-cultural models that, by citation count, appear to garner the highest levels of research attention. We chose to include (1) the GLOBE project’s nine practices dimensions, (2) GLOBE’s nine values dimensions (House et al., 2004), (3) Hofstede’s (2010) six cultural dimensions, (4) Ronen and Shenkar’s (2013) 11 cultural clusters, and (5) Schwartz Value Survey’s (2006, 2008) seven dimensions (see Table 2 for all dimension names). Finally, as a geographic region approach, we chose to include (6) 14 of the 17 United Nations M49 sub-regions. (Three sub-regions contained no findings for the analyzed 136 bivariate relationships: Melanesia, Micronesia, Polynesia.) Thus, in total, we have analyzed six cultural models containing a sum of 56 individual dimensions.

Table 1 Uncorrected summary estimates and number of effects contributing to 136 bivariate relationships

Construct	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Conscientiousness		1,225	1,241	200	112	619	83	36	104	54	35	23	52	819	235	173	217
2. Extraversion	.176		1,298	263	140	701	58	34	121	38	24	36	37	597	177	217	101
3. Emo. stability	.232	.199		319	153	582	121	42	152	38	44	18	20	497	118	180	163
4. Negative states	-.122	-.075	-.320		1,581	440	2,209	923	1,349	453	371	81	506	1,134	386	336	956
5. Positive states	.187	.210	.202	-.181		238	341	122	383	165	95	58	69	721	289	174	205
6. Ability	.076	.083	.111	-.094	.129		87	39	117	75	42	30	49	3,488	138	476	93
7. Stressors	-.061	-.062	-.178	.252	-.058	-.046		893	1,154	429	115	55	292	501	221	308	437
8. Autonomy	.122	.165	.135	-.126	.220	.114	-.058		591	144	77	24	136	236	146	79	114
9. Job satisfaction	.172	.148	.177	-.279	.335	.173	-.143	.317		1,258	301	144	888	805	388	459	537
10. Org. commitment	.125	.109	.095	-.152	.354	.130	-.110	.272	.483		361	118	836	606	559	428	211
11. Justice/fairness	.076	.176	.095	-.134	.260	.037	-.202	.162	.384	.360		58	181	206	493	152	298
12. LMX	.132	.145	.038	-.156	.295	.114	-.025	.310	.397	.363	.483		67	256	153	323	25
13. Turnover intent	-.083	-.104	-.152	.280	-.170	-.113	.212	-.191	-.437	-.400	-.295	-.295		240	192	147	238
14. Role behav.	.147	.092	.081	-.038	.131	.175	-.016	.116	.160	.164	.171	.265	-.083		1,130	1,232	248
15. Extra-role behav.	.188	.132	.066	-.057	.234	.162	.009	.166	.225	.211	.191	.265	-.158	.418		466	408
16. Leadership	.100	.154	.059	-.071	.282	.147	-.014	.188	.275	.253	.267	.366	-.151	.204	.228		194
17. CWB	-.195	-.020	-.127	.250	-.124	.026	.235	-.097	-.193	-.151	-.155	-.034	.239	-.095	-.146	-.059	

Values below the diagonal represent multilevel (effects nested by sample) meta-analytic summaries corrected for sampling error. Values above the diagonal represent numbers of effect sizes (the mean number of participants per sample was 382).

Org. organizational, Behav. behavior, LMX leader-member exchange, CWB counterproductive work behavior.

Several differences between the models are worth noting. First, two approaches – Ronen and Shenkar and M49 – are based on a bucketing or “containers” approach (Kirkman et al., 2017: 19). These approaches place countries into categories and, thus, statistical analyses of these distinctions are based on categorical predictors. Ronen and Shenkar (2013, figure 4) arranged 96 countries into 11 clusters that take into account the ecocultural factors of religion and language (Georgas & Berry, 1995). This approach yields clusters of countries on different continents (e.g., the cultural cluster labeled “Anglo” is comprised of Australia, Canada, Ireland, New Zealand, United Kingdom, and United States). In contrast, the UN M49 Standard categorizes countries into five contiguous regions (Africa, Americas, Asia, Europe, Oceania) and then 17 sub-regions (e.g., Northern America) and “does not imply any assumption regarding political or other affiliation” (United Nations Statistics Division, 1999).

In contrast to the “containers”-based approaches, the remaining four approaches explain cultural phenomena using continuous dimensions. Perhaps best known is Hofstede’s (1980; see also Hofstede et al., 2010) six-dimension framework (individualism, indulgence, long-term orientation, masculinity, power distance, and uncertainty avoidance). Similarly, the Schwartz Value Survey (2006, 2008) contains seven continuous dimensions (harmony, embeddedness, hierarchy, mastery, affective autonomy, intellectual autonomy, and egalitarianism). Finally, the GLOBE project (House et al., 2004) contains nine dimensions (uncertainty avoidance, future orientation, power distance, institutional collectivism, humane orientation, performance orientation, in-group collectivism, gender egalitarianism, and assertiveness), which are assessed in two distinct ways to ascertain values and practices.

Thus, our selection of varied cross-cultural models and 136 common bivariate relationships is purposely designed to afford a 30,000-foot view of culture’s moderating effect in applied psychology. To summarize, two of our models use a “container” approach and four use a multidimensional or multifactorial approach. Within the container approaches, the Ronen and Shenkar (2013) model is explicitly intended to index culture, and the other (i.e., U.N. M49) is explicitly intended to index geography. Within the four multidimensional approaches, three are intended to index values and one is intended to index practices. We chose these diverse models to allow many substantive

comparisons. For example, cross-cultural researchers should be very interested to learn whether the Ronen and Shenkar (2013) container approach explains more variance in findings than the M49 container approach. To our knowledge, we are unaware of a large-scale investigation of such questions, especially in the context of moderating effects.

Analytic Approach

As initial steps, we filtered the metaBUS database to include only individual-level data (e.g., group- and organization-level data were excluded), leaving 854,022 rows of data, or roughly 85% of the database. (Each row represents one published correlation coefficient and associated metadata.) We created new database fields and populated them with country-to-container mappings using the Ronen and Shenkar (2013) and M49 (United Nations Statistics Division, 1999) classifications. Finally, we located dimensions scores by country for GLOBE (House et al., 2004), Hofstede et al. (2010), and the Schwartz Value Survey (1992, 2006), and mapped the country-level values to our database. In cases of reported antonyms (e.g., job dissatisfaction, neuroticism), correlations were reverse-scored.

Estimation of incremental variance explained

The first and primary goal of our analyses was to estimate the incremental variance in 136 meta-analytic databases’ findings attributable to each cultural model. To this end, we conducted series of multilevel meta-analyses (MLM) with unreliability-corrected effect sizes as the dependent variable and studied inverse variance as the weighting factor. In our case, effect sizes are nested in samples, which are in turn nested in countries (e.g., Rockstuhl et al., 2020). For a review of three- to five-level meta-analyses, see Fernández-Castilla et al. (2020); for an example analogous to our data structure with R script as supplemental material, see Lehtonen et al. (2018). [We also report results based on non-nested weighted least squares analyses in Supplemental Materials (SM) Table 4; their results yield generally similar conclusions.]

The key outcome of interest is the change in adjusted pseudo- R^2 comparing two models: Model 1 contains publication year as the only predictor of corrected effect size, and Model 2 contains publication year and the cultural model (e.g., six Hofstede dimensions). Using this approach, we account for a variety of statistical artifacts (i.e., sampling

error, unreliability, potential temporal trends) and ascertain the incremental variance accounted for by each cultural model by entering all their dimensions simultaneously. (We also conduct rounds of analyses on the 56 individual dimensions separately.)

According to Schmidt and Hunter (2015), the majority of artifactual variance in summary estimates is explained by sampling error and unreliability. Thus, although the metaBUS database contains other information that could be leveraged as additional control variables, such as participant sample type, we sought to limit model complexity for more stable parameter estimates. (In the metaBUS database, sample type is indexed using two categorical variables – one for the source of the information and another for the target, resulting in a potentially highly complex control variable, see Bosco et al., 2017.) Furthermore, we make our R script data files available at (<http://osf.io/32c4y>) for further exploration.

In all, we conducted 25,296 MLM analyses (i.e., 136 bivariate relationships $\times$ [56 individual cultural dimensions + 6 omnibus cultural dimensions] $\times$ 3 multilevel models) using the `rma.mv` function in the metafor R package (Viechtbauer, 2010). Specifically, for each of the 136 meta-analytic datasets, we extracted the key outcome (i.e., change in adjusted pseudo- R^2 due to culture) iteratively for each of the six models, requiring 816 MLM analyses. As an example of analyses in this round, this represents the pseudo- R^2 change attributable to all six Hofstede dimensions, entered simultaneously, on job satisfaction-in-role performance effects. Next, we conducted a similar round of analyses at the level of the individual dimensions (e.g., pseudo- R^2 change attributable to Hofstede's individualism dimension on job satisfaction-in-role performance effects), necessitating an additional 7616 MLM analyses (i.e., 136 bivariate relationship datasets crossed with a total 56 dimensions).

We summarize the key outcome (i.e., adjusted pseudo- R^2 change due to culture) in two ways. First, we collapse across bivariate relationship to obtain and compare k_{Effects} -weighted mean estimates of pseudo- R^2 for each model (e.g., incremental variance due to Hofstede's dimensions vs. incremental variance due to GLOBE's). Next, we collapse across the six cross-cultural models to investigate the extent to which the 136 bivariate relationships vary in their sensitivity to cultural moderation effects. Finally, as a robustness assessment, we investigate

the extent to which the culture moderation effect for the 136 bivariate relationships is explained by sampling error.

Estimation of mean effect size difference

A secondary goal of our study was to estimate the typical observed moderation effect size attributable to culture in applied psychology research. Thus, in these analyses, we correct for sampling error, but not for unreliability or publication year because uncorrected effects serve as the basis for power analyses. Put differently, our goal in this subset of analyses is to ascertain the extent to which *observed* effects vary across cross-cultural models and their dimensions.

To this end, we rerun 7,616 of the MLMs on uncorrected effect sizes using the nesting structure described in the previous section. For each of these analyses, we estimate a single moderating effect representing either a categorical (e.g., Anglo = 1, non-Anglo = 0) or continuous (e.g., individualism) cultural dimension. From each of these analyses, we extract the standardized slope coefficient for continuous cultural predictors, or the unstandardized slope coefficient for categorical cultural predictors. Thus, each continuous moderator's effect is interpreted as a Pearson's r , and each categorical moderator's effect is interpreted as a Cohen's q . Note that, according to Cohen (1988), q and r are interpreted in the same way. To arrive at omnibus summaries across the individual dimensions, we weight the r and q values by k_{Effects} . As an example, the weighted mean of the six r values pertaining to Hofstede's six dimensions represents the mean effect size difference due to Hofstede's model.

The summarization of the key outcome (i.e., $|q|$ for categorical cultural predictors and $|r|$ for continuous) mirrors the procedure outlined for the incremental variance analyses.

RESULTS

Overview

We organize our results according to our two main goals. In the first section, we report the incremental variance results and report the relative performance of (i.e., variance explained by) the six cross-cultural models (e.g., Hofstede vs. GLOBE), collapsed across (i.e., k_{Effects} -weighted by) the 136 bivariate relationships. (Mean uncorrected effects and their number of effects are provided in Table 1.) Next, we report

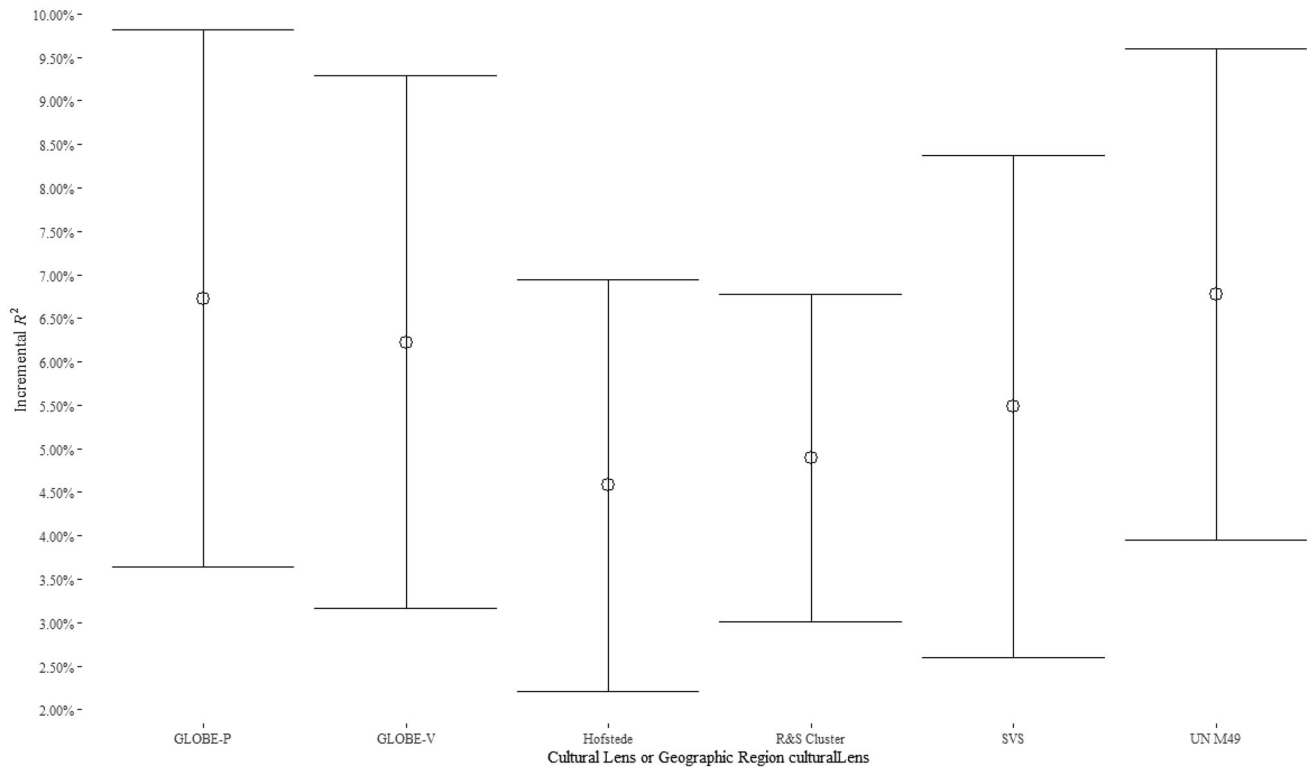


Figure 1 Mean of incremental adjusted pseudo- R^2 values due to culture or region for 136 common bivariate relationships. Mean values are weighted by k (i.e., the number of effects contributing to each meta-analytic summary). Confidence intervals (95%) are based on the standard error of the difference comparing the two adjusted R^2 values.

the incremental variance results by specific cross-cultural dimensions (e.g., Hofstede's individualism vs. GLOBE's individualism), again collapsed across the 136 bivariate relationships. Finally, we report variance among the 136 bivariate relationships and report robustness assessments to ascertain whether the variance among the specific bivariate relationships is due to real culture effects or, alternatively, sampling error.

The second portion of our results follows a similar format, with the key outcome being the magnitude of observed (i.e., uncorrected) moderating effects.

Incremental Variance Due to Culture

As described above, incremental variance analyses were conducted using MLM with unreliability-corrected correlations as the dependent variable, publication year as a control variable, inverse variance as the weighting factor, and cross-cultural dimensions as predictors (i.e., moderators).

When collapsed across the 136 meta-analytic datasets and across the six cross-cultural models, the overall k_{Effects} -weighted mean value for

incremental adjusted pseudo- R^2 attributable to cross-cultural model is roughly 6%. Incremental adjusted pseudo- R^2 with all dimensions entered simultaneously, explained by cultural lenses increased from Hofstede's dimensions (Δ adjusted $R^2 = 4.6\%$) to Ronan and Shenkar's cultural clusters (Δ adjusted $R^2 = 4.9\%$), the Schwartz Values Survey (SVS; Δ adjusted $R^2 = 5.5\%$), GLOBE's values dimensions (Δ adjusted $R^2 = 6.2\%$), GLOBE's practices dimensions (Δ adjusted $R^2 = 6.7\%$), and UN's M49 Standard (Δ adjusted $R^2 = 6.8\%$). However, as shown in Figure 1, the 95% confidence intervals based on the standard error of the difference (i.e., adjusted pseudo- R^2 Model 1 vs. adjusted pseudo- R^2 Model 2) overlap substantially. (Note that we employed a conservative approach to estimating the confidence interval based on the standard error of the difference, which ultimately relies on $n = 136$ meta-analytic summaries.)

Regarding specific cross-cultural dimensions (i.e., within models), as shown in Table 2, incremental adjusted R^2 varied across dimensions when entered individually into the MLM analyses. As an example, collapsed across the 136 bivariate relationships,

Table 2 Incremental variance explained and mean effect size differences by cultural lens or geographic region container for 136 common bivariate relationships

Geographic cluster or cultural dimension	<i>k</i> Effects	Δ Adj. R^2	Mean $ r $ or $ q $
GLOBE practices dimensions (aggregate)	48,687	.067 ^a	.064 ^b
Uncertainty avoidance	48,687	.005	.057
Future orientation	48,687	.013	.064
Power distance	48,687	.013	.064
Institutional collectivism	48,687	.009	.065
Humane orientation	48,687	.011	.066
Performance orientation	48,687	.012	.061
In-group collectivism	48,687	.015	.073
Gender egalitarianism	48,687	.010	.057
Assertiveness	48,687	.010	.069
GLOBE values dimensions (aggregate)	48,687	.062 ^a	.063 ^b
Uncertainty avoidance	48,687	.013	.075
Future orientation	48,687	.006	.052
Power distance	48,687	.011	.066
Institutional collectivism	48,687	.017	.075
Humane orientation	48,687	.007	.057
Performance orientation	48,687	.008	.063
In-group collectivism	48,687	.005	.053
Gender egalitarianism	48,687	.014	.066
Assertiveness	48,687	.010	.063
Hofstede cultural dimensions (aggregate)	49,397	.046 ^a	.068 ^b
Individualism	50,027	.014	.077
Indulgence	49,491	.013	.073
Long-term orientation	50,118	.008	.061
Masculinity	50,027	.007	.064
Power distance	50,027	.012	.065
Uncertainty avoidance	50,027	.014	.070
Ronen and Shenkar cultural cluster (aggregate)	50,166	.049 ^a	.047 ^b
Africa	20	.003	.133
Anglo	35,768	.014	.040
Arab	153	.016	.131
Confucian	3829	.007	.048
East Europe	625	.007	.107
Far East	738	.006	.078
Germanic	2564	.008	.069
Latin America	142	.004	.164
Latin Europe	2879	.012	.066
Near East	581	.007	.095
Nordic	3808	.007	.057
Schwartz value survey dimensions (aggregate)	50,073	.055 ^a	.068 ^b
Harmony	50,073	.010	.075
Embeddedness	50,073	.008	.064
Hierarchy	50,073	.008	.057
Mastery	50,073	.011	.071
Affective autonomy	50,073	.011	.069
Intellectual autonomy	50,073	.011	.078
Egalitarianism	50,073	.010	.064
United Nations M49 geographic region (aggregate)	50,294	.068 ^a	.047 ^b
Australia and New Zealand	2327	.008	.065
Central Asia	1	.010	.267
Eastern Asia	3185	.007	.055
Eastern Europe	420	.006	.113
Latin America and the Caribbean	137	.005	.156
Northern Africa	7	.048	.401

Table 2 (Continued)

Geographic cluster or cultural dimension	k_{Effects}	$\Delta \text{Adj. } R^2$	Mean $ r $ or $ q $
Northern America	31,204	.011	.036
Northern Europe	3242	.012	.073
Southeastern Asia	741	.007	.087
Southern Asia	446	.006	.099
Southern Europe	1213	.019	.083
Sub Saharan Africa	46	.003	.135
Western Asia	981	.012	.088
Western Europe	6344	.009	.050

k_{Effects} = Number of effect sizes; $\Delta \text{Adj. } R^2$ = Incremental adjusted R^2 for unreliability-corrected effects attributable to region or culture beyond publication year (k -weighted mean of 136 bivariate relationships). Mean $|r|$ or $|q|$: mean association between cultural dimension and uncorrected effect size ($|r|$), or cultural cluster and uncorrected effect size ($|q|$) across 136 bivariate relationships.

^a Aggregate adjusted R^2 values are based on regression models containing all dimensions or clusters simultaneously.

^b Aggregate mean $|r|$ or $|q|$ values represent the weighted mean of the absolute differences ($|q|$) or correlations ($|r|$) for dimension and cluster relationships, respectively.

Hofstede's individualism (when entered alone) incremented adjusted R^2 by 1.4% compared with long-term orientation and masculinity, which explained roughly half that value (i.e., 0.8% and 0.7%, respectively). Similarly, GLOBE's in-group collectivism (practices dimension) outperformed its uncertainty avoidance dimension by a factor of three (i.e., 1.5% and 0.5%, respectively). In contrast, for GLOBE's values dimensions, the opposite pattern was observed (i.e., in-group collectivism = 0.5%, uncertainty avoidance = 1.3%). Finally, the SVS dimensions exhibited less variance in the adjusted R^2 change (range = 0.8–1.1%). Thus, within models, some cross-cultural dimensions explain variance in findings nearly twice as well as others. Furthermore, although GLOBE's practices and values models perform similarly overall (i.e., when all dimensions are entered simultaneously), the R^2 values for the individual dimensions within each model seem to show opposite patterns (i.e., $r_s = -.78$).

A more nuanced investigation of the incremental R^2 values for each of the 136 bivariate relationships revealed that some bivariate relationships are more sensitive to cross-cultural effects (collapsed across models). As examples, job satisfaction–counterproductive behaviors and ability–conscientiousness relationships present with relatively strong cross-cultural moderation effects (see Figure 2). In contrast, job satisfaction–turnover intention relationships present with minimal cultural moderation effects.

Importantly, the number of effects summarized appears to vary with bivariate relationship type, which makes a robustness assessment appropriate. In this case, the question addressed is whether the

apparent differences across bivariate relationships are due to cross-cultural effects or, alternatively, artifactual variance (e.g., sampling error). As shown in Figure 3, our results indicate the presence of clear “funnel” patterns between k_{Effects} and mean incremental R^2 for each of the six cross-cultural models. Put differently, larger R^2 values are associated with smaller meta-analytic databases, and more precise estimates (i.e., larger datasets) present with smaller R^2 values. Thus, as can be seen in Figure 3, much of the variance ostensibly attributable to bivariate relationship type is attributable to sampling error.

Analyzing datasets with few countries can yield multicollinear country-level moderator variables. Thus, as another robustness analysis, we limit our analyses to cases in which 10 or more countries contributed findings (see SM Table 3). In general, the R^2 values were slightly weaker in the limited set. However, the pattern remained similar, and the adjusted R^2 values reported in Table 2 and their analogues in the supplemental materials correlate at $r = .99$.

Mean Observed Moderating Effects Due to Culture

As described above, estimation of moderating effects due to culture was carried out using effect sizes uncorrected for unreliability and without statistical control for attributes such as publication year. Thus, findings in this section refer to *observed* effect sizes, and may be used as input to power analysis for the purpose of estimating required sample sizes for future investigation. Overall, the mean moderating effect due to culture was roughly $|r|$ or $|q| = .06$ (see Table 2). As an illustrative example of this magnitude, imagine a study with two correlation matrices, one based on

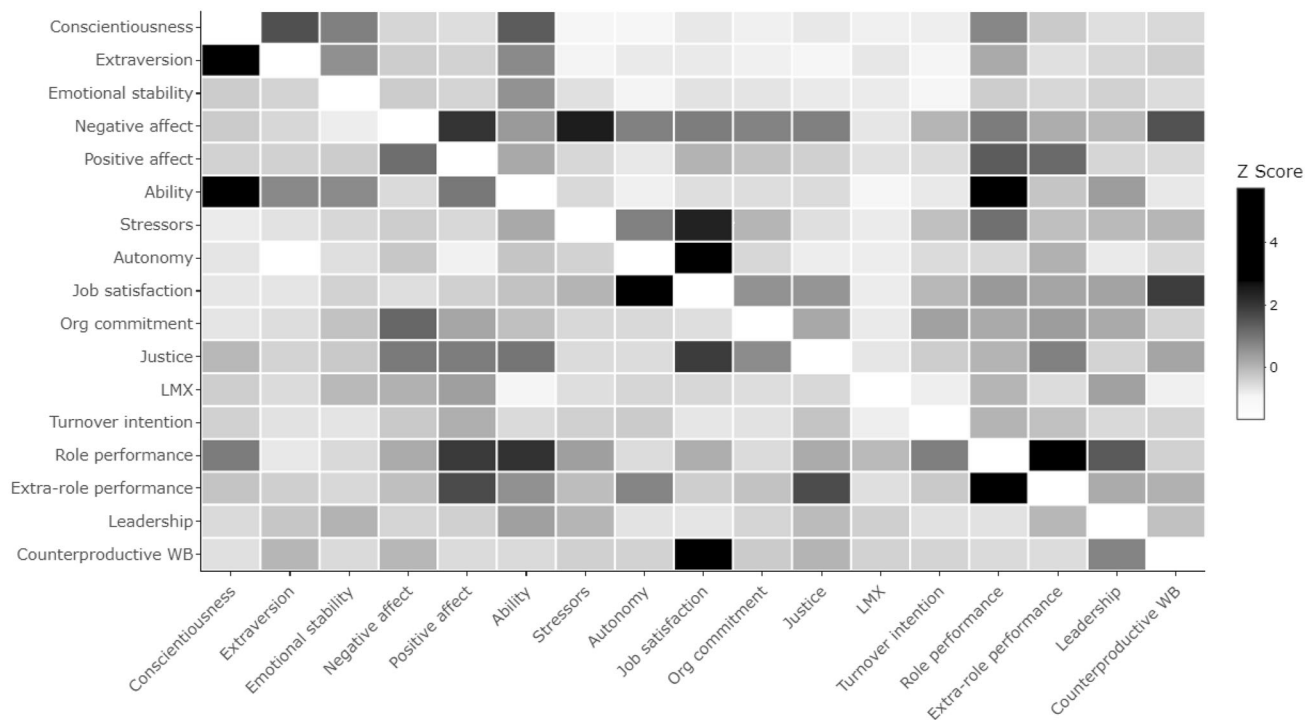


Figure 2 Mean culture effect for 136 bivariate relationships expressed as incremental adjusted pseudo- R^2 (below diagonal) and mean effect size difference (i.e., $|r|$ or $|q|$; above diagonal). Cell shading represents the influence Z score (i.e., effect multiplied by k , which penalizes summary estimates with low k).

Anglo data and another based on Confucian data. For a given bivariate relationship (e.g., negative states – turnover intention), $z_t = .31$ in the Confucian sample and $z_t = .25$ in the Anglo sample (or vice-versa), resulting in $|q| = .06$, an effect size requiring a combined sample size of 8,728 to reach the conventional 80% level of statistical power (Faul et al., 2009). (We revisit statistical power in the Discussion.)

Collapsed across the 136 bivariate relationships, k_{Effects} -weighted moderating effect sizes attributable to individual cultural dimensions presented some similarities with the incremental variance results. As shown in Table 2, like the R^2 change results, Hofstede's individualism presented with the largest moderating effects ($|r|_M = .077$) and long-term orientation the weakest ($|r|_M = .061$). Similarly, like the R^2 change results, GLOBE practices' in-group collectivism presented with the largest moderating effects ($|r|_M = .073$) and uncertainty avoidance the weakest ($|r|_M = .057$). Likewise, within the GLOBE values model, institutional collectivism moderating effects were largest ($|r|_M = .075$) and those for future orientation ($|r|_M = .052$) and in-group collectivism ($|r|_M = .053$) were weakest. Like our incremental variance analyses, we again observed an unexpected negative

relationship in effect size differences due to GLOBE's practices and values dimensions ($r = -.56$).

Unlike the incremental variance analyses, for which we were able to estimate all cultural dimensions simultaneously or individually, for the mean moderating effects we simply took the k_{Effects} -weighted mean of the individual models' moderating effects. Weighted moderating effects increased from Ronen and Shenkar and UN M49 ($|q|_M = .047$) to GLOBE values ($|r|_M = .063$), followed by GLOBE practices ($|r|_M = .064$) and SVS and Hofstede ($|r|_M = .068$). Thus, the two "container" or clustering models presented with the smallest mean moderating effects and values- and practices-based models with the largest.

As carried out for the incremental variance analyses, we investigated mean moderating effects for each of the 136 bivariate relationships in search of patterns. We also conducted a similar set of robustness analyses to investigate the extent to which apparent bivariate relationship sensitivity effects are explained by sampling error. As shown in Figure 2 (above diagonal), some bivariate relationships present with larger cross-cultural moderating effects. Similar to the R^2 results, job satisfaction–counterproductive behaviors and ability–conscientiousness relationships present with

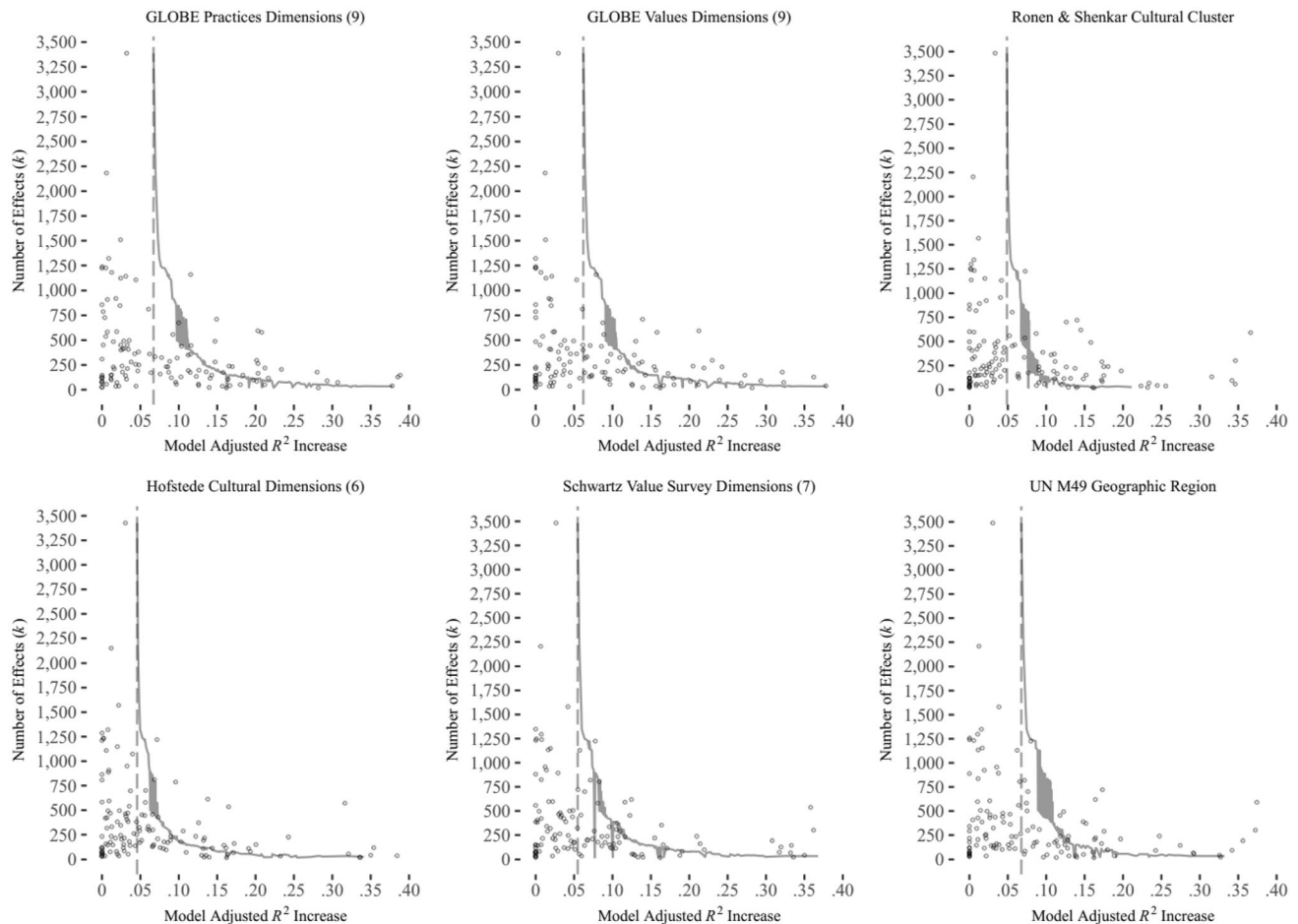


Figure 3 Scatter plots of adjusted pseudo- R^2 increase due to cultural lens or geographic region container against k (i.e., number of effects) for 136 meta-analytic summaries. The solid line represents a cumulative, k -weighted summary estimate. The dashed line represents the final cumulative mean (i.e., overall k -weighted mean).

relatively strong cross-cultural moderation effects. In addition, like the R^2 change results, job satisfaction–turnover intention relationships present with relatively small $|r|$ and $|q|$ values. However, as shown in Figure 4, again, clear “funnel” shapes are evident, indicating the presence of sampling error (i.e., small summaries presented with larger effects, and larger summaries with smaller effects).

DISCUSSION

We set out to benchmark approximate values for culture’s impact (e.g., incremental variance explained) according to six popular cultural theoretic perspectives. We focus exclusively on culture-as-moderator effects which, we suspect, should be the effect type of greater interest to applied psychologists. As an example of this reasoning, consider a researcher studying the relationship

between co-worker incivility and employee turnover. The researcher may wish to float a hypothesis regarding cross-cultural moderation (e.g., from individualism) and then conduct a study. At this point, for study planning purposes (e.g., power analysis), existing culture-as-main effect findings are not useful because they address a completely different question. As an example, the main effect between individualism and turnover, while interesting, would only provide information on whether the overall rate of turnover tends to be greater in more individualistic cultures. Indeed, even if the researcher were supplied with both zero-order main effects (i.e., correlations for individualism–incivility and individualism–turnover), they would be unable to estimate the moderating effect of individualism on the bivariate relationship of interest. That is, the researcher would not know whether *culture matters* to the relationship under study. Thus, when one

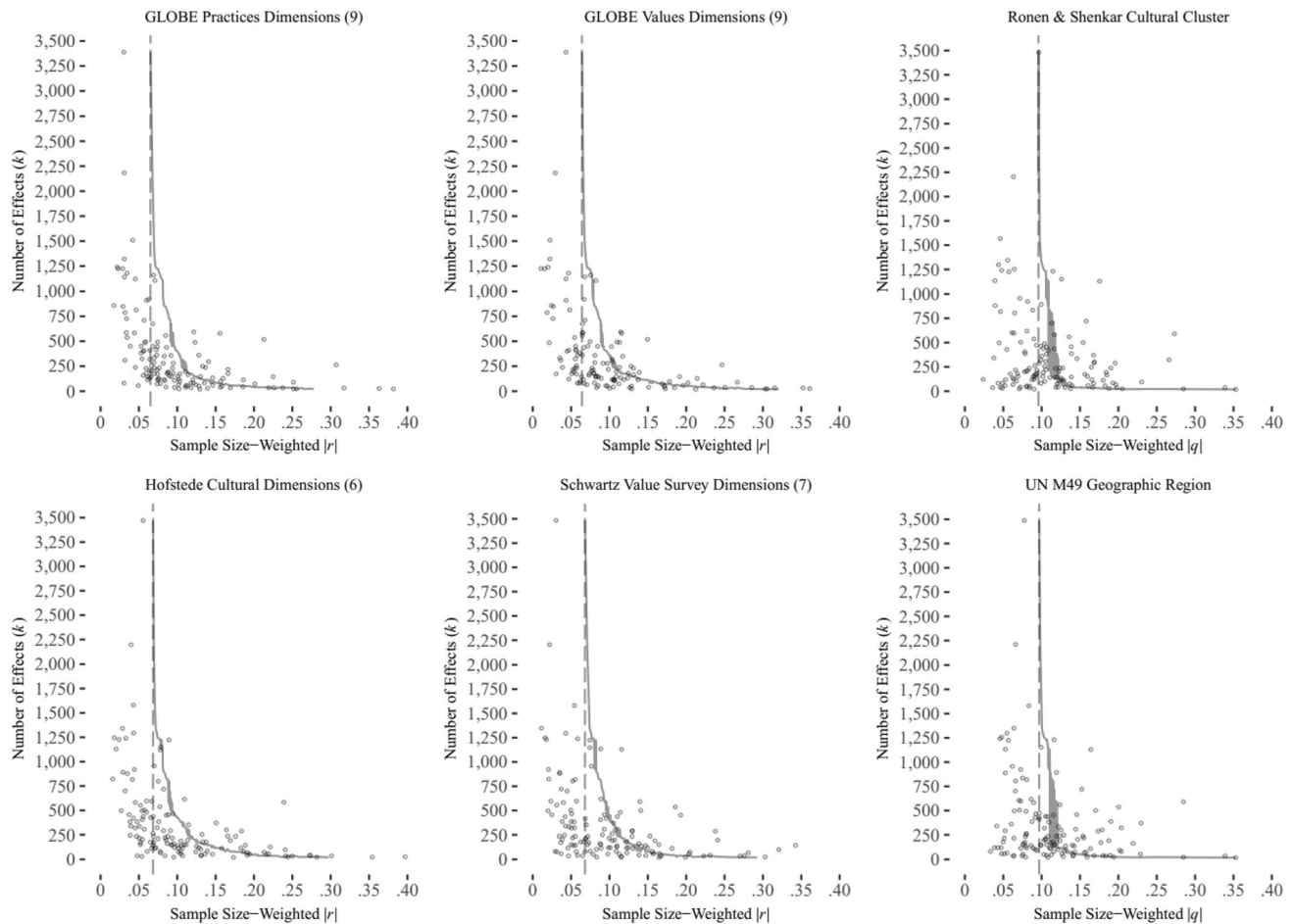


Figure 4 Scatter plots of multilevel mean $|r|$ (for continuous cultural predictors) or $|q|$ (for categorical cultural/regional predictors) against k (i.e., number of effects) for 136 meta-analytic summaries. The solid line represents a cumulative, k -weighted summary estimate. The dashed line represents the final cumulative mean (i.e., overall k -weighted mean).

thinks about it, the moderating effect should be of greater interest because it describes whether culture influences bivariate relationships (or more complex phenomena), which are the building blocks of theory. Our summary of culture-as-moderator effects is the first of its kind (cf. Taras et al., 2010a) and has several important implications for research and practice.

While our results did not provide a strong case for the superiority of one cultural perspective over any other, our results contribute to the literature by providing a rough benchmark for culture's impact in applied psychology. Put plainly, that impact is roughly 6% when all of a given model's dimensions are entered simultaneously, a value similar to incremental variance explained by including multiple simultaneous moderators in substantive organizational research. As one recent meta-analytic example, Ellen III et al. (in press) observed

incremental variance due to the Dark Triad personality traits above and beyond the Big Five of 5.2% and 8% for organizational and interpersonal deviance, respectively. Thus, culture-as-moderator effect sizes are on a par with substantive moderators in applied psychology when multiple moderators are tested simultaneously.

However, in many cases, not all cultural dimensions are included in a given study. For example, it is common to examine the moderating effect due to individualism alone. Using our data, the analogous variance explained for individual dimensions is 1.1%. To compare this value to an existing benchmark, Aguinis et al.'s (2005) content analysis of hypothesized moderating effects (excluding gender and ethnicity) was $f^2 = .013$ (equal to 1.3% ΔR^2 ; see Aguinis et al., Table 1). Thus, our relatively coarse analyses (i.e., 56 individual cross-cultural dimensions crossed with 136 commonly

investigated bivariate relationships) indicate that, taken together, the effect of culture is similar to the effect of hypothesis-relevant moderators in organizational research when estimated individually. In fact, several of the individual cultural dimensions reached values greater than Aguinis et al.'s (2005) mean (e.g., Hofstede's individualism = 1.4%), meaning that some cultural dimensions matter *more* than the typical substantive moderators native to organizational research.

Our results reveal that cultural models tend to vary in explanatory power, although the comparisons did not reach statistical significance. In these analyses, perhaps the most surprising finding is that the M49 model explained more variance than Ronen and Shenkar's. This is surprising because the latter reflects three ecocultural factors (i.e., geography, religion, and language) and, thus, should be expected to index culture more than the M49 classification system, which reflects only geography. As one possible explanation, our Ronen and Shenkar findings may represent a restricted variance interaction (Cortina et al., 2019). That is, the moderation effect may be attenuated by a restriction of variance caused by homogeneity of religion and language in Ronen and Shenkar's clusters. Still, it is surprising that the explicit efforts to model culture did not perform better than simple contiguous geographic boundaries.

Regarding individual dimensions within cultural models, we observed that some dimensions performed two or three times as well as others (e.g., Hofstede's individualism outperformed masculinity by a factor of two). In other cases, dimensions performed similarly across cultural models. As examples, the individualism dimension was the top performer for both Hofstede's model and GLOBE's practices model, and power distance explained roughly the same variance whether measured by GLOBE's practices, GLOBE's values, or Hofstede's model. In still other cases, we observed unexpected trends. Most notably, the R^2 values for GLOBE practices' and GLOBE values' dimensions were negatively correlated ($r_s = .78$). That is, when a given GLOBE dimension performed well as a practice, it tended to perform poorly as a value, and vice versa. As one possible explanation for this finding, existing research indicates a negative correlation between the GLOBE practice and values dimensions with national-level data (Hofstede, 2006; Maseland & van Hoorn, 2009; Taras, Steel, & Kirkman, 2010b).

Robustness analyses revealed the usually unsurprising 'funnel' pattern of effect size against sample size (i.e., increasing variance in effects at smaller sample sizes). However, what was surprising was that not until k_{Effects} reached 500 to 750 did the R^2 or observed effect size differences stabilize. Importantly, consider that, in a typical meta-analysis of zero-order effects, one might expect stabilization to occur by $k = 50$. However, our study summarizes moderating effects, which tend to be less stable. An alternative explanation for the funnel shape is that weaker effects are studied more frequently; however, we do not expect this to be case. For researchers, these findings indicate that presumed moderating effects due to culture – especially when atypically large in magnitude – might, in fact, be attributable to sampling error. Furthermore, our findings indicate that massive corpora of existing findings are necessary to estimate moderating effect magnitudes attributable to culture with any certainty.

When compared to the incremental variance findings, those pertaining to our second research question (i.e., observed mean effect size differences) revealed some similarities and differences. (The correlation between the individual dimensions' R^2 values and mean differences presented in Table 2 was $r = .56$.) First, overall, mean effect size differences ranged from .05 to .07. We remind the reader that our comparison of two effects represents a moderation effect. As an example, the relationship between job satisfaction and in-role performance might be $z_r = .15$ in an Anglo culture and $z_r = .21$ in a Confucian culture. This difference, $q = .06$, is typical of our mean observed differences across 136 bivariate relationships and through multiple cultural models.

Mean observed effect size differences were weakest for M49 (i.e., geographic regions) and Ronen and Shenkar's (2013) model (i.e., country clusters). All other models yielded larger effect size differences that, like our incremental variance analyses, were indistinguishable from one another. (We find this pattern much easier to explain than the pattern observed in the incremental variance analyses.) Furthermore, like the incremental variance analyses, the GLOBE practices and values dimensions were again negatively correlated ($r = -.56$). However, unlike the incremental variance findings, we observed that mean differences for individual dimensions presented with much narrower ranges. As an example, the strongest Hofstede dimension (i.e., individualism, $|r| = .077$) was only 26% larger

in magnitude than the weakest dimension (i.e., long-term orientation, $|r| = .061$).

Perhaps the most important implication of our mean differences findings is the unusually large sample sizes that would be required to achieve conventional levels of statistical power; we revisit this in the following section.

Implications for Research and Practice

Four important implications for research and practice can be delineated from our results. The first implication pertains to how culture may be conceptualized in future studies. Our results indicate that Hofstede's (1980) values model and Ronen and Shenkar's (2013) configurational model generally explained the least variance in 136 bivariate relationships. Indeed, this is a surprising observation given that Ronen and Shenkar's cultural mapping process was intended to go beyond geography (e.g., Northern America; M49) and act as a substitute for the controversial cultural values (e.g., individualism; Hofstede, 1980). Again, it is possible that the phyletic approach they used only captured the diversity and homogeneity of countries along relatively objective stable ecocultural factors (e.g., language, religion), not the nuanced factors that potentially contribute to their respective cultures. Importantly, we are not recommending that future researchers abandon these approaches. However, our results may indicate that the efficacy of ecocultural profiles that integrate both relatively stable (e.g., geography) and polycontextual (e.g., local economy, welfare socialism) factors for explaining cross-cultural effects should be explored. Indeed, a better knowledge of how stable and polycontextual ecocultural factors interact to explain cross-cultural effects should lead to more informed evidence-based practice recommendations, which will likely help international business practitioners to create cultural-specific organizational practices and, thus, narrow the science–practice gap.

More recent cross-cultural models tend to have the greatest explanatory power. This is noteworthy because it indicates that cross-cultural researchers have improved how they conceptualize and measure culture over time (Aguinis et al., 2020). Although our results suggest that recently developed and updated meta-theoretical perspectives on culture may help researchers best explain how culture moderates bivariate relationships, we do not support the culling of traditional perspectives (e.g., Hofstede, 1980). Indeed, our findings indicate

that, even when models perform relatively poorly as a whole, they may contain specific individual dimensions with high explanatory potential.

The second major implication of this study pertains to statistical power in cross-cultural research. Specifying an informed target effect size presents a challenge to conducting a priori power analysis (Cohen, 1992). Ideally, cross-cultural researchers would rely on an effect size that accounts for both emic (i.e., culture-specific; Gelfand et al., 2002) and substantive (i.e., bivariate relationship) factors rather than taking “a shot in the dark with Cohen's (1988) benchmarks” (Bosco et al., 2015: 441). Supplemental Materials Tables 1 and 2 report benchmarks for culture-as-moderator effects (in terms of variance due to culture) and cross-cultural effect size difference benchmarks (in terms of Cohen's [1988] q and r) for 136 commonly investigated relationships (e.g., job satisfaction–turnover intention) for each specific cultural model and dimension. Indeed, the distribution of these effects suggests that existing benchmarks for “small,” “medium,” and “large” area differences (e.g., Cohen's 1988 $|q| = .1, .3, \text{ and } .5$, respectively) are unrealistically high and do not generalize to the cross-cultural literature.

Given that Cohen's (1988) effect size benchmarks for r and q are comparable, it is not unreasonable to assume that Bosco et al.'s (2015) correlation coefficient benchmarks can be adapted to q values. Yet, even under this assumption, Bosco et al.'s benchmark for a medium effect (i.e., $|q| = .16$) applied to culture-as-moderator effects is noticeably larger than what was observed in our study. As such, a researcher who relies on Bosco et al.'s (2015) benchmarks will assume that a total of 1,234 participants are needed to detect a medium cross-cultural difference. However, if the “true” cross-cultural difference is, on average, $|q| = .010$ (see SM Table 2), this sample size will achieve only 42% statistical power (indeed, only 8% statistical power is achieved when $n = 356$, which is the suggested total sample size needed to detect a medium cross-cultural difference per Cohen's (1988) benchmarks). As such, the use of Bosco et al.'s (2015) and Cohen's (1988) benchmarks – which do not account for cross-cultural effects – may lead to underpowered cross-area studies (Maxwell, 2004). An inspection of SM Table 2 suggests that “small,” “medium,” and “large” cross-cultural differences should be operationalized as $|q| = .066, .086, \text{ and } .11$, respectively, which represents the 33rd, 50th, and 67th percentiles of the distribution of the 816

culture-as-moderator effect sizes observed in our study. We contend that these cross-cultural differences benchmarks will help researchers to design studies that test cross-cultural theory with adequate statistical power. To facilitate this endeavor, we introduce CASST (Cross-Area Sample Size Tool; see <https://casst.shinyapps.io/gen1/>), an open-access software that allows users to interact with our results, and can be used by cross-cultural researchers to improve *a priori* sample size estimation.

A third implication of our results pertains to the interpretation of cross-cultural research findings and the estimation of practical significance. Imagine that a researcher or human resources practitioner of a multinational firm encounters recent uncorrected meta-analytic mean effect size estimates for the relationship between agreeableness and affective commitment (AC) in collectivistic ($k = 10$, $\bar{r} = .35$) and individualistic ($k = 19$, $\bar{r} = .19$) cultures (Choi et al., 2015). Although it seems evident that agreeableness explains more variance in AC in collectivistic cultures than individualistic cultures, there exist no benchmarks to characterize the magnitude of the moderation in the context of existing cross-cultural research. The empirically derived benchmarks reported in our study suggest that the effect (i.e., $q = .173$) represents a relatively large (i.e., 88th percentile) cross-cultural moderation effect size.

A fourth implication of this study regards the statistical power of meta-analytic studies on cross-cultural differences. Figures 3 and 4 offer insight into the effect of sampling error on the observed culture-as-moderator effects. A half-funnel is observed in each panel – indicating that more extreme cross-cultural differences are observed in meta-analytic summaries with small k_{Effects} . To the best of our knowledge, there exists no empirically derived statistical power recommendations for the meta-analytic studies on cross-cultural differences. This is surprising given that there is a danger of drawing incorrect conclusions if a meta-analysis is performed too early (i.e., before enough studies are available; Thorlund et al., 2011). Our results indicate that meta-analytic studies on cross-cultural differences in substantive phenomena with fewer than 500 effects should be interpreted with caution. Indeed, such large k_{Effects} values are likely required to detect culture-as-moderator effects given that moderator effects are typically very small.

Limitations and Future Directions

Although our findings are consistent with previous cautions regarding the prevalence and magnitude of moderating effects (e.g., Aguinis et al., 2005), a number of limitations must be addressed. First, we concede that the culture-as-moderator effects observed in the current study may be due to some artifact not controlled for in the current study. For example, variance in culture-as-moderator effects due to response style (e.g., acquiescence), which can distort cross-cultural results (see Cheung & Rensvold, 2000), is not accounted for. This is fertile ground for future research.

A second limitation of our study is that we summarized culture-as-moderator effects by using data drawn from the metaBUS database (Bosco et al., 2017). Although the volume and variety of this corpus of data is impressive, the metaBUS database currently has curated data from journals printed in English only. Therefore, it remains possible that including data from other sources or unpublished findings might influence our conclusions. Relatedly, an inspection of our results indicates that a large portion of the metaBUS database is likely comprised of “WEIRD” samples (e.g., 62% of effect sizes included in the M49 analyses originated from Northern America). Indeed, research findings typically focus almost exclusively on a single cultural context, the United States, despite it sharing only 5% of the world’s population (Thalmayer et al. 2021). Thus, the prevalence of “WEIRD” samples in the metaBUS database is not a feature of the platform itself, but rather a reflection of the relative homogeneity of samples in the available literature. Still, it is important to note that an uneven distribution of moderator variables lowers the statistical power to detect moderator effects (Hempel et al., 2013), which may suggest that certain results reported in the current study – particularly those that deviate noticeably from a 50:50 ratio in the moderator variable distribution and have relatively small sample sizes – should be interpreted with caution. However, we are encouraged by the fact that cross-cultural research is growing over time (Gelfand et al., 2017).

A third limitation is that we did not conduct analyses for publication bias; we chose this approach for three reasons. First, publication bias analyses are typically conducted on zero-order effects extracted from primary sources (e.g., publications, unpublished sources). The logic follows that some portion of the effect sizes are unavailable

for summary. However, our analyses do not rely on those sorts of effects. Put differently, we derive moderating effect size magnitudes (e.g., q) from meta-analytic summaries (e.g., mean r). Thus, the effect sizes we summarize were never the scrutinized in the review process. Second, there exists disagreement among publication bias researchers regarding which methods to employ and how to interpret the magnitude of publication bias and the extent to which it threatens the validity of published results. Third, reporting these analyses would be arduous, as we conducted thousands of meta-analyses. Thus, we urge the reader to keep in mind what might be called second-order publication bias and consider our results generalizable to published research. Still, we make available our research materials (e.g., datasets, analytic scripts) to scholars interested in examining how publication bias may threaten the results of the current study (see <http://osf.io/32c4y>).

A fourth limitation of our study is that only 136 applied psychology relationships were used to summarize culture-as-moderator effects. The magnitude of moderating effects may be different in bivariate relationships not summarized here, or in other scientific disciplines. Still, our corpus is comparatively comprehensive. Drawing on a database of 598 studies containing 1,507 effects Taras et al. (2010a; see Table 5) reported a total of 22 benchmarks for relationships between cultural dimensions (e.g., individualism) and five broad outcomes (e.g., job performance). In the current study, we report 816 culture-as-moderator effect size benchmarks (i.e., 136 bivariate relationships crossed with six cultural models) from more than 50,000 findings. Therefore, although we do not present an exhaustive review of the entire literature, we feel confident that the current study adds to our cumulative knowledge on moderating effect sizes attributable to culture. Still, we encourage scholars to conduct similar analyses using bivariate relationships not included in our study, as well as in future cross-cultural studies, and in other scientific spaces (e.g., strategic management, marketing) so that the scientific community can take a more holistic account of the prevalence and magnitude of culture-as-moderator effects.

Finally, several review articles (e.g., Gelfand et al., 2017; Kirkman et al., 2017) have outlined in great detail numerous future directions for cross-cultural researchers. For example, Kirkman et al. (2006) provided a large-scale review of cross-cultural research that used Hofstede's dimensions, and

offered three recommendations for future cross-cultural research, (1) to more explicitly consider and report effect size so as to find out not *whether* culture matters, but also *how much* it matters; (2) to more carefully consider mappings of countries to groups; and (3) to explore dimensions beyond those offered by Hofstede (1980). In the present study, we shed light on each of these persistent uncertainties. Still, we recognize that no single study can address the population of recommendations for future cross-cultural research. However, we encourage cross-cultural scholars to address the "grand challenges" (Buckley et al., 2017: p. 1046) facing them and to tackle the proposed research agendas outlined in recent review articles.

As additional future research directions, research could examine other cross-cultural models such as Trompenaars's model of national culture differences (Trompenaars & Woolliams, 2004) or Gelfand et al.'s (2006) cultural tightness-looseness dimension. Future research could also consider adding statistical control for additional methodological or cultural factors (e.g., Rockstuhl et al., 2020). As another option to at least partially ameliorate the concerns related to our reliance on metaBUS, future research could investigate similar effects in published meta-analyses which, if conducted alongside a proper systematic review, should be less influenced by missing studies or publication bias.

However, we expect that the most fruitful avenues for rapid progress come in the form of alternative analytic approaches. For example, apart from understanding the overall R^2 attributable to each model, it would be beneficial to conduct comparisons to determine whether models provide incremental variance beyond each other (e.g., does GLOBE explain variance in effects after controlling for Hofstede?). Relatedly, we had aimed to provide relative importance analyses that would compare the contribution to R^2 for all cultural models' dimensions entered simultaneously. However, there exists a great deal of natural multicollinearity between the moderator levels, and we were unable to overcome that limitation using ridge regression. Ultimately, we relied on multilevel meta-analysis, which necessitated the use of a high-performance computing cluster due to computational complexity and the number of iterations. Another option is to use robust variance estimation rather than multilevel meta-analysis (Hedges, Tipton, & Johnson, 2010), which is less resource-intensive. As a final possible future research question, we would be interested to learn how the results would change

had we relied instead on a dichotomization of continuous cultural dimensions (e.g., individualism as 0 or 1; Rockstuhl, 2020), which appears common in the cross-cultural literature despite the availability of statistical procedures that handle continuous inputs.

CONCLUSION

We have observed that incremental variance due to culture is roughly 6% depending on the perspective employed (all dimensions estimated simultaneously) or 1% (individual cultural dimensions) across 136 bivariate relationships common to applied psychology. This moderation effect size information is important for cross-cultural researchers to calibrate the answer to how much culture matters. Furthermore, we observed typical mean effect size differences between .05 and .07 for $|r|$ or $|q|$, which are interpreted similarly. This information is important for study planning to ensure that sufficient statistical power is achieved. Finally, we observed differences in the extent to which bivariate relationships exhibit cross-cultural effects. However, we also observed that moderating effects due to culture tend to stabilize only with impressively large corpora (i.e., over 500 effects). Thus, cross-cultural researchers should be aware of the possibility that cultural moderation effects are at least partially attributable to sampling error. Finally, our study offers information that can be used by practitioners to evaluate the relative effectiveness and generalizability of substantive findings across multiple cultures.

REFERENCES

- Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. 2005. Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology*, 90(1): 94–107.
- Aguinis, H., Cascio, W. F., & Ramani, R. S. 2017. Science's reproducibility and replicability crisis: International business is not immune. *Journal of International Business Studies*, 48(6): 653–663.
- Aguinis, H., Ramani, R. S., & Cascio, W. F. 2020. Methodological practices in international business research: An after-action review of challenges and solutions. *Journal of International Business Studies*, 51(9): 1593–1608.
- Allen, T. D., French, K. A., Dumani, S., & Shockley, K. M. 2015. Meta-analysis of work–family conflict mean differences: Does national context matter? *Journal of Vocational Behavior*, 90: 90–100.
- Beugelsdijk, S., Kostova, T., & Roth, K. 2017. An overview of Hofstede-inspired country-level culture research in international business since 2006. *Journal of International Business Studies*, 48(1): 30–47.
- Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. 2015. Correlational effect size benchmarks. *Journal of Applied Psychology*, 100(2): 431–449.
- Bosco, F. A., Uggerslev, K. L., & Steel, P. 2017. MetaBUS as a vehicle for facilitating meta-analysis. *Human Resource Management Review*, 27(1): 237–254.
- Buckley, P. J., Doh, J. P., & Benischke, M. H. 2017. Towards a renaissance in international business research? Big questions, grand challenges, and the future of IB scholarship. *Journal of International Business Studies*, 48(9): 1045–1064.
- Busse, C., Kach, A. P., & Wagner, S. M. 2017. Boundary conditions: What they are, how to explore them, why we need them, and when to consider them. *Organizational Research Methods*, 20(4): 574–609.
- Cascio, W. F., & Aguinis, H. 2008. Staffing Twenty-first-century Organizations. *Academy of Management Annals*, 2(1): 133–165.
- Cheung, G. W., & Rensvold, R. B. 2000. Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, 31(2): 187–212.

ACKNOWLEDGEMENTS

We thank Dr. Bo Nielsen and two anonymous reviewers for highly constructive feedback that allowed us to substantially improve our manuscript. We also thank Dr. Piers Steel for reviewing an earlier version of this manuscript and Dr. Brad Price for his support. This research was supported by the Super Computing System (Thorny Flat) at West Virginia University, which is funded in part by the National Science Foundation (NSF) Major Research Instrumentation Program (MRI) Award #1726534. An earlier version of this manuscript was presented at the 34th Annual Meeting of the Society for Industrial and Organizational Psychology.

NOTES

¹On average, each M49 sub-region is comprised of 14.59 countries (SD = 14.83). However, two sub-regions are made up of a relatively large number of individual countries (i.e., Sub-Saharan Africa [$n_{\text{countries}} = 53$] and Latin America and the Caribbean [$n_{\text{countries}} = 52$] and, thus, may be viewed as potential outliers. After removing these two sub-regions, each remaining sub-region has an average of 9.53 countries (SD = 4.32), which still exceeds the recommended number of countries needed for the detection of cross-cultural effects (Franke & Richey, 2010).

- Choi, D., Oh, I.-S., & Colbert, A. E. 2015. Understanding organizational commitment: A meta-analytic examination of the roles of the five-factor model of personality and culture. *Journal of Applied Psychology*, 100(5): 1542–1567.
- Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. Erlbaum.
- Cohen, J. 1992. A power primer. *Psychological Bulletin*, 112(1): 155–159.
- Cortina, J. M., Koehler, T., Keeler, K. R., & Nielsen, B. B. 2019. Restricted variance interaction effects: What they are and why they are your friends. *Journal of Management*, 45(7): 2779–2806.
- Crede, M., Jong, J., & Harms, P. 2019. The generalizability of transformational leadership across cultures: a meta-analysis. *Journal of Managerial Psychology*, 34(3): 139–155.
- Ellen III, B. P., Alexander, K. C., Mackey, J. D., McAllister, C. P., & Carson, J. E. in press. Portrait of a workplace deviant: A clearer picture of the Big Five and Dark Triad as predictors of workplace deviance. *Journal of Applied Psychology*.
- 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(4): 1149–1160.
- Fernández-Castilla, B., Jamshidi, L., Declercq, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. 2020. The application of meta-analytic (multi-level) models with multiple random effects: A systematic review. *Behavior Research Methods*, 52(5): 2031–2052.
- Fisher, D. M. 2014. A multilevel cross-cultural examination of role overload and organizational commitment: Investigating the interactive effects of context. *Journal of Applied Psychology*, 99(4): 723–736.
- Franke, G. R., & Richey, R. G. 2010. Improving generalizations from multi-country comparisons in international business research. *Journal of International Business Studies*, 41(8): 1275–1293.
- Gelfand, M. J., Aycan, Z., Erez, M., & Leung, K. 2017. Cross-cultural industrial organizational psychology and organizational behavior: A hundred-year journey. *Journal of Applied Psychology*, 102(3): 514–529.
- Gelfand, M. J., Nishii, L. H., & Raver, J. L. 2006. On the nature and importance of cultural tightness-looseness. *Journal of Applied Psychology*, 91(6): 1225–1244.
- Gelfand, M. J., Raver, J. L., & Holcombe Ehrhart, K. 2002. Methodological issues in cross-cultural organizational research. In S. Rogelberg (Ed.), *Handbook of industrial and organizational psychology research methods*. Oxford: Blackwell.
- Georgas, J., & Berry, J. W. 1995. An ecocultural taxonomy for cross-cultural psychology. *Cross-Cultural Research*, 29(2): 121–157.
- Hedges, L. V., Tipton, E., & Johnson, M. C. 2010. Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1): 39–65.
- Hempel, S., Miles, J. N. V., Booth, M. J., Wang, Z., Morton, S. C., & Shekelle, P. G. 2013. Risk of bias: A simulation study of power to detect study-level moderator effects in meta-analysis. *Systematic reviews*, 2: 107–207.
- Hofstede, G. 1980. *Culture's consequences: International differences in work-related values*. Sage.
- Hofstede, G. 2006. What did GLOBE really measure? Researchers' minds versus respondents' minds. *Journal of International Business Studies*, 37(6): 882–896.
- Hofstede, G., Hofstede, G. J., & Minkov, M. 2010. *Cultures and organizations: Software of the mind* (3rd ed.). McGraw-Hill.
- House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. 2004. *Culture, leadership, and organizations*. Thousand Oaks: Sage.
- Kashima, Y., Bain, P. G., & Perfors, A. 2019. The psychology of cultural dynamics: What is it, what do we know, and what is yet to be known? *Annual Review of Psychology*, 70(1): 499–529.
- Kirkman, B. L., Lowe, K. B., & Gibson, C. B. 2006. A quarter century of culture's consequences: A review of empirical research incorporating Hofstede's cultural values framework. *Journal of International Business Studies*, 37(3): 285–320.
- Kirkman, B. L., Lowe, K. B., & Gibson, C. B. 2017. A retrospective on culture's consequences: The 35-year journey. *Journal of International Business Studies*, 48(1): 12–29.
- Knafo, A., Roccas, S., & Sagiv, L. 2011. The value of values in cross-cultural research: A special issue in honor of Shalom Schwartz. *Journal of Cross-Cultural Psychology*, 42(2): 178–185.
- Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. 2018. Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin*, 144(4): 394–425.
- Maseland, R., & van Hoorn, A. 2009. Explaining the negative correlation between values and practices: A note on the Hofstede–GLOBE debate. *Journal of International Business Studies*, 40(3): 527–532.
- Matsumoto, D. 2006. Culture and cultural worldviews: Do verbal descriptions about culture reflect anything other than verbal descriptions of culture? *Culture and Psychology*, 12(1): 33–62.
- Maxwell, S. E. 2004. The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods*, 9(2): 147–163.
- Minkov, M. 2018. A revision of Hofstede's model of national culture: old evidence and new data from 56 countries. *Cross Cultural and Strategic Management*, 25(2): 231–256.
- Platt, J. R. 1964. Strong Inference. *Science*, 146(3642): 347–353.
- Ramesh, A., & Gelfand, M. J. 2010. Will they stay or will they go? The role of job embeddedness in predicting turnover in individualistic and collectivistic cultures. *Journal of Applied Psychology*, 95(5): 807.
- Rockstuhl, T., Eisenberger, R., Shore, L. M., Kurtessis, J. N., Ford, M. T., Buffardi, L. C., & Mesdaghinia, S. 2020. Perceived organizational support (POS) across 54 nations: A cross-cultural meta-analysis of POS effects. *Journal of International Business Studies*, 51(6): 933–962.
- Ronen, S., & Shenkar, O. 2013. Mapping world cultures: Cluster formation, sources and implications. *Journal of International Business Studies*, 44(9): 867–897.
- Schmidt, F. L., & Hunter, J. E. 2015. *Methods of meta-analysis: Correcting error and bias in research findings* (3rd edn.). Thousand Oaks: Sage.
- Schwartz, S. H. 1992. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology*, 25(1): 1–65.
- Schwartz, S. H. 2006. A theory of cultural value orientations: Explication and applications. *Comparative Sociology*, 5(2–3): 137–182.
- Schwartz, S. H. 2008. The 7 Schwartz cultural value orientation scores for 80 countries.
- Segall, M. H., Lonner, W. J., & Berry, J. W. 1998. Cross-cultural psychology as a scholarly discipline: On the flowering of culture in behavioral research. *American Psychologist*, 53(10): 1101–1110.
- Shao, R., Rupp, D. E., Skarlicki, D. P., & Jones, K. S. 2013. Employee justice across cultures: A meta-analytic review. *Journal of Management*, 39(1): 263–301.
- United Nations Statistics Division. 1999. Standard Country or Area Codes for Statistics Use, 1999 (Revision 4).
- Taras, V., Kirkman, B. L., & Steel, P. 2010a. 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.
- Taras, V., Steel, P., & Kirkman, B. L. 2010b. Negative practice-value correlations in the GLOBE data: Unexpected findings, questionnaire limitations and research directions. *Journal of International Business Studies*, 41(8): 1330–1338.
- Taras, V., Steel, P., & Kirkman, B. L. 2016. Does country equate with culture? Beyond geography in the search for cultural



- boundaries. *Management International Review*, 56(4): 455–487.
- Teagarden, M. B., Von Glinow, M. A., & Mellahi, K. 2018. Contextualizing international business research: Enhancing rigor and relevance. *Journal of World Business*, 53(3): 303–306.
- Thalmayer, A. G., Toscanelli, C., & Arnett, J. J. 2021. The neglected 95% revisited: Is American psychology becoming less American? *American Psychologist*, 76(1): 116–129.
- Thorlund, K., Imberger, G., Walsh, M., Chu, R., Gluud, C., Wetterslev, J., Guyatt, G., Devereaux, P. J., & Thabane, L. 2011. The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis – A simulation study. *PLoS ONE*, 6(10): e25491.
- Trompenaars, F., & Woolliams, P. 2004. *Business across cultures*. New York: Wiley.
- Tsui, A. S., Nifadkar, S. S., & Ou, A. Y. 2007. Cross-national, cross-cultural organizational behavior research: Advances, gaps, and recommendations. *Journal of Management*, 33(3): 426–478.
- Tung, R. L., & Verbeke, A. 2010. Beyond Hofstede and GLOBE: Improving the quality of cross-cultural research. *Journal of International Business Studies*, 41(8): 1259–1274.
- Viechtbauer, W. 2010. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, 1(3): 2010.
- Zhong, L., Wayne, S. J., & Liden, R. C. 2016. Job engagement, perceived organizational support, high-performance human resource practices, and cultural value orientations: A cross-level investigation. *Journal of Organizational Behavior*, 37(6): 823–844.

ABOUT THE AUTHORS

James Field is an Assistant Professor of Management in the John Chambers College of Business and Economics at West Virginia University, USA. He earned his PhD at Virginia Commonwealth University. His program of research focuses on organizational research methods and open science topics.

Frank Bosco is a faculty member at the Department of Management and Entrepreneurship at

Virginia Commonwealth University, USA. He is Director of the metaBUS.org project, and conducts research on various topics in management, applied psychology, and meta-science. He is particularly interested in approaches for summarizing, visualizing, and making openly available large-scale landscapes of scientific evidence.

David Kraichy is an Assistant Professor of Human Resources and Organizational Behaviour at the Edwards School of Business, University of Saskatchewan, Canada. His current research interests center on issues relating to talent management.

Krista Uggerslev is the Applied Research Chair in Leadership and Talent at the Northern Alberta Institute of Technology, Canada. Her work focuses on solving the real-world challenges facing employers and governments, particularly around talent management and labor shortages. Along with Drs. Bosco and Field, Dr. Uggerslev contributes to the metaBUS project, which provides cloud-based research synthesis atop the world's largest curated database of social science research findings.

Mingang K. Geiger is a PhD Candidate at the John Chambers College of Business and Economics, West Virginia University, USA. Her primary research explores implicit biases that disadvantage marginalized groups in the workplace. She is also interested in how big data can contribute to a better understanding of organizational phenomena.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Accepted by Bo Nielsen, Consulting Editor, 11 March 2021. This article has been with the authors for two revisions.