Journal of International Business Studies (2010) 41, 1581–1588

www.jibs.net

RESEARCH NOTE

Effect sizes and the interpretation of research results in international business

Paul D Ellis

Department of Management & Marketing, Hong Kong Polytechnic University, Kowloon, Hong Kong

Correspondence:
PD Ellis, Department of Management & Marketing, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.
Tel: +1852 2766 7108;
Fax: +1852 2765 0611

Abstract

Journal editors and academy presidents are increasingly calling on researchers to evaluate the substantive, as opposed to the statistical, significance of their results. To measure the extent to which these calls have been heeded, I aggregated the meta-analytically derived effect size estimates obtained from 965 individual samples. I then surveyed 204 studies published in the *Journal of International Business Studies*. I found that the average effect size in international business research is small, and that most published studies lack the statistical power to detect such effects reliably. I also found that many authors confuse statistical with substantive significance when interpreting their research results. These practices have likely led to unacceptably high Type II error rates and invalid inferences regarding real-world effects. By emphasizing p values over their effect size estimates, researchers are under-selling their results and settling for contributions that are less than what they really have to offer. In view of this, I offer four recommendations for improving research and reporting practices.

Journal of International Business Studies (2010) **41,** 1581–1588.
doi:10.1057/jibs.2010.39

Keywords: evaluation of current empirical approaches; theory–method intersection; meta-analysis; statistical power; effect size

INTRODUCTION

A fundamental goal of international business research is to estimate the magnitude and direction of effects that exist in the international business domain. Effects are commonly expressed in terms of measures of association between two or more variables (the so-called *r*-family of effects), or the degree of difference between two or more groups (the *d*-family of effects) (Ellis, 2010b).

International business researchers typically estimate population effects by examining representative samples. Although researchers may invest considerable effort in minimizing measurement and sampling error, and thereby producing more accurate effect size estimates, ultimately the goal is a better understanding of population effects. This distinction between population effects and researchers' estimates of those effects is critical to understanding the difference between substantive and statistical significance. *Statistical significance* reflects the improbability of findings drawn from samples, given certain assumptions about the null hypothesis. *Substantive significance* is concerned with meaning, as in, what do the findings say about population effects themselves?

Received: 7 August 2009
Revised: 13 July 2010
Accepted: 25 July 2010

For some time journal editors and academy presidents have been calling on authors to evaluate the substantive significance of their results (Campbell, 1982; Rynes, 2007; Shaver, 2006). These calls are framed in appeals for research that matters (Hambrick, 1994), and which generates knowledge that is "relevant and useful to practitioners" (Cummings, 2007: 357). An important takeaway from this is that researchers should explicitly report and interpret their estimates of the effect size (Iacobucci, 2005; Shaver, 2008; Zedeck, 2003).

The calls for meaningful interpretation based on estimates of effect size have been consistent and clear, but have they been heeded? To answer this question I read more than 200 empirical studies recently published in the *Journal of International Business Studies* (*JIBS*). To get some sense of the scale of the effects that international business researchers study, I also conducted a census of meta-analyses published in 32 international business journals. My goal was to answer three related questions:

(1) What is a typical effect size in international business research?
(2) Does the average study have sufficient statistical power to detect effects of this size?
(3) To what extent are international business researchers interpreting the substantive significance of their results?

After addressing these questions I outline four recommendations for improving research practice.

EFFECT SIZES IN INTERNATIONAL BUSINESS

To identify typical effect sizes in international business research I used the electronic database ProQuest to identify every meta-analysis published in 32 international business journals between 1995 and the summer of 2009. (Full details of the methods used are available from the author.) As a result of my search I was able to identify 23 weighted mean effect sizes. These weighted means were originally obtained by pooling a total of 965 individual study-specific estimates drawn from an aggregate sample of N=223,800.

The most interesting thing about the mean effect sizes was that none of them was particularly big. With reference to Cohen's (1988: 79–80) thresholds for defining small, medium and large effects in the correlational metric (i.e., $r \geqslant 0.10$, 0.30 and 0.50 respectively), five of the mean effects were small (i.e., $0.10 \leqslant r < 0.30$), and 18 were smaller than small (i.e., $r < 0.10$). The weighted mean of the 23 weighted mean effect sizes obtained from the meta-analyses was just r=0.06. In percentage of variance terms this means that the average international business effect is equivalent to less than half of one percent.

The miniature size of international business effects is in itself not particularly noteworthy. Meta-analyses done in other disciplines routinely reveal effects that are small in size (e.g., Mazen, Graf, Kellogg, & Hemmasi, 1987). In view of these realities, the onus is on researchers to ensure that their studies are sufficiently empowered to detect small effects.

THE POWER OF INTERNATIONAL BUSINESS RESEARCH

The power of any test of statistical significance is defined as the probability that it will reject a false null hypothesis. In other words, statistical power is the likelihood that a study will detect an effect when there is an effect there to be detected. By assessing the statistical power of sets of studies, power analysts can quantify the average likelihood of Type II errors in published research (Cashen & Geiger, 2004; Mazen et al., 1987). Statistical power directly affects the quality of inferences drawn from samples. If the average statistical power of research is low, the proportion of results that are either inconclusive or incorrect will inevitably be high.

Brock (2003) provided the first assessment of statistical power in the domain of international business. His assessment was based on research published in the period from 1990 to 1999. Brock found that the majority of international business studies lacked statistical power. In view of this, Brock recommended that researchers address statistical power and the attendant threat of Type II errors before committing to projects.

Curious to see whether Brock's (2003) study has affected research practice, I surveyed every empirical study published in *JIBS* from 2003 to 2008. My aim was to conduct a census of all studies that relied on tests of statistical significance for inference-making. To be included in the analysis a study needed to report sample sizes and the results of tests of statistical significance. In the 39 issues covered by the review period there were 189 empirical articles reporting the results of 204 independent studies.

As there is little value in running retrospective power analyses based on effect sizes observed within individual samples (Hoenig & Heisey, 2001), I followed past practice and calculated the

Table 1 Statistical power in *JIBS* (2003–2008)

Power	*Small effect size*		*Medium effect size*		*Large effect size*	
	Frequencies	*Cumulative %*	*Frequencies*	*Cumulative %*	*Frequencies*	*Cumulative %*
(A) All studies (N=204)						
0.80–1.00	31	15	182	89	202	99
0.70–0.79	4	17	6	92	0	99
0.60–0.69	11	23	3	94	0	99
0.50–0.59	10	27	4	96	1	100
<0.50	148	100	9	100	1	100
Mean	0.48		0.93		0.99	
Median	0.36		0.99		0.99	
SD	0.32		0.15		0.06	
(B) Non-panel data (N=151)						
0.80–1.00	18	12	132	87	149	99
0.50–0.79	23	27	10	94	1	99
<0.50	114	100	9	100	1	100
Brock's mean[a]	0.33		0.81		0.95	
Mean	0.38		0.91		0.99	
Median	0.28		0.99		0.99	
SD	0.25		0.16		0.07	

[a]Based on studies published in *JIBS* from 1990–1999.
Note: Frequencies above the dotted line are the number of studies achieving conventional levels of power (0.80).

statistical power for a range of three hypothetical effect sizes. (Full details of the methods used are available from the author.) The summary results of the power survey are presented in Panel A of Table 1. The mean power to detect small, medium and large effects was found to be 0.48, 0.93 and 0.99 respectively. This means that in research settings where effects were there to be detected, the probability of committing a Type II error for the average study was 52% (or 1–0.48) if population effects were small, 7% if they were medium-sized, and just 1% if they were large. If a study with sufficient power is defined as one that has at least an 80% chance of detecting legitimate effects (Cohen 1988), then the majority of studies published in *JIBS* (85%) lacked the power needed to detect small effects.

In my reading I noticed an increasing trend towards the use of panel data sets. Panel data are multidimensional in the sense that observations encompass multiple cases observed over multiple time-periods. As observations are clustered by firm, panel data are to some degree non-independent, making it difficult to compute statistical power (Abraham & Russell, 2008). For this reason sampling decisions are commonly framed as an optimization process based on rules of thumb (Scherbaum & Ferreter, 2009). Given these limitations, it was not possible to calculate accurate power levels for the 53 studies in the sample that relied on panel data sets. Consequently I recalculated mean power levels based on the 151 studies that did not make use of panel data, and the results are shown in Panel B of Table 1. Given that there were relatively few published studies using panel data during Brock's (2003) review, the findings summarized in Panel B are the best for determining whether there has been any increase in statistical power between the two review periods. The mean power to detect small, medium and large effects was found to be 0.38, 0.91 and 0.99 respectively, in contrast with Brock's corresponding figures of 0.33, 0.81 and 0.95. In terms of the number of studies attaining adequate power levels, the results of my analysis reveal that only 12% of studies were sufficiently empowered to have a reasonable chance of detecting effects of any size. This is only slightly higher than Brock's comparable figure of 10%, suggesting that any increase in the power of studies has been marginal at best.

The results of my power survey reinforce Brock's original conclusions. Although the average statistical power of studies published in *JIBS* remains high in comparison with research published in other leading business journals (Cashen & Geiger 2004), the majority of studies (up to 88%) continue to lack the power required for the detection of small effects. This has adverse consequences for drawing inferences about population effects, as I will discuss below.

RESEARCH PRACTICES IN INTERNATIONAL BUSINESS

The results of my power survey reveal that the majority of studies published in *JIBS* lacked the statistical power to detect small effects. In a discipline where average effect sizes are small, inattention to matters of statistical power can lead to both Type I and Type II errors (Ellis, 2010b). To minimize the threat of these errors, it is essential that studies are sufficiently empowered to detect sought-after effects, and that researchers distinguish statistical significance testing from the estimation and interpretation of those effects. If international business researchers are sensitive to these issues, this would be apparent in the design of studies and the interpretation of research results.

The Design of International Business Research

If studies are designed with effect sizes in mind, reviews of past research would focus on effect size estimates obtained by others, and researchers would target minimum sample sizes based on their prior expectations regarding the effect being studied. In my reading of published research it was clear that the majority of *JIBS* authors either ignored or made no explicit reference to these issues when designing studies. In well over half of the studies I read, target sample sizes were set with no explicit rationale. In about a third of the studies the sample size was determined by data or resource availability, and in others N was justified with reference to past practice. In only a handful of studies were desired sample sizes explicitly informed by statistical power concerns (e.g., Barden, Steensma, & Lyles, 2005). I also found a few instances where authors referred to the power implications of their sample sizes when reporting or interpreting the results of tests of statistical significance (e.g., Child, Chung, & Davies, 2003; Parboteeah, Hoegl, & Cullen, 2008). As the design of most studies was uninformed by prior expectations regarding effect sizes and the attendant requirements for statistical power, it was not difficult to find studies that were either under- or over-powered.

The authors of one underpowered study (N=203) hypothesized that a firm's global technological competence would be positively related to its international market orientation. However, their test result for this effect generated a standardized coefficient of 0.13 that was judged to be statistically nonsignificant. Yet if we assume that an international market orientation has an effect on technological competence equivalent to r=0.13, this study would have needed an additional 259 observations for this result to have had a reasonable chance (defined as β=0.80) of achieving statistical significance with α set at the conventional level of 0.05. If we hypothesized a directional effect, permitting a less stringent one-tailed test, an additional 161 observations would have been required. As the sample size used in this study was well below both of these minima, we cannot rule out insufficient power as a probable factor accounting for this statistically nonsignificant result.

In an underpowered study genuine effects can go undetected, but in an overpowered study everything is statistically significant. The challenge then becomes one of separating meaningful effects from trivial ones. This was the challenge that confronted the authors of a large-N analysis of the effects of market liberalization in China. These authors hypothesized that industrial goods manufacturers would perform well as a consequence of support provided by the central government. Based on their analysis of more than 100,000 firm-year observations, and the statistical significance of their result, they concluded that their hypothesis was supported. But given the massive power of this study, the statistical significance of any particular test result was never in doubt. Everything was statistically significant. A better way to gauge the effect of industrial policy would have been to examine the effect size estimates directly. These were very small, just −0.03 for the raw correlation and −0.02 for the standardized regression coefficient when performance was measured in terms of profitability. Any effect China's industrial policies had on performance was essentially negligible, equivalent to between 1/11th and 1/25th of 1% in percentage of variance terms.

The Interpretation of Research Results

Are international business researchers in the habit of reporting and interpreting estimates of effect size? In many of the studies I read it was apparent that researchers confused statistical with substantive significance when interpreting their results. This usually happened when conclusions about effects were drawn solely by looking at the p values of test results. The problem with this is that the p values generated by statistical significance tests are confounded indices that reflect both the size of the effect as it occurs in the population and the sample size used to detect it. As N goes up p goes down, irrespective of the underlying effect size. The implication is that trivial results are sometimes interpreted as meaningful in large-N studies, while meaningful results are sometimes written off as "nonsignificant" in small-N studies.

By emphasizing p values over effect sizes, researchers are under-selling their results. They are throwing away hard-earned data, and settling for contributions that are less than what they really have to offer. An example of missing the bigger contribution is provided by my own study of the effects of trade on market orientation (Ellis, 2007). In Table 4 of that study I compared the market orientation of two groups of firms. As predicted, the 74 firms that were highly dependent on distant markets were found to have lower market orientation scores than the 56 firms that were essentially domestic marketers. Satisfied that this statistically significant difference confirmed my expectations regarding the adverse effects of selling to distant markets, I then moved on to my next statistical test. What I did not assess at that time was the magnitude of the difference between the two groups, and whether this difference might actually be meaningful to exporters.

Using the data provided in Table 4 of Ellis (2007), I have since calculated that the standardized mean difference between the two groups was 0.88 – that is, close to one standard deviation, and a large difference by Cohen's (1988) standards. In a field dominated by small and trivial-sized effects a large difference should attract attention, yet I missed it! Firms that export to distant and diverse markets will have a much lower market orientation than rivals selling to neighboring and similar markets. Given that market orientation accounts for as much as 12% in the variation in firm performance (Cano, Carrillat, & Jaramillo, 2004), a difference of this magnitude would also be of substantial interest to practitioners.

In many of the studies I read, effect sizes were reported unintentionally. This occurred whenever authors reported test statistics that happen to double as effect size indicators (e.g., r, R^2, η^2). However, on only a few occasions were these results were explicitly identified as estimates of the underlying effect size (e.g., Ellis, 2008; Luk et al., 2008). In one study a distinction was made between the statistical significance of a result and the "substantive importance" implied by the proportion of variance accounted for (Child et al., 2003). On other occasions effects were interpreted, but without a meaningful frame of reference. Some authors mentioned that their models "performed well", while others observed that their combination of predictors generated R^2s (or an increase in R^2s) that were "respectable" or "remarkable". But without a context such claims are meaningless. Performed well with respect to what? Respectable in comparison with what? Equally meaningless was the claim that a test statistic was bigger or more impressive than a result obtained in an earlier study. If separate studies are estimating the same population effect, then results should be celebrated for converging rather than diverging. And if they are estimating different effects, then there is little to be gained by comparing them.

Reading these studies, it soon became clear that the majority of international business researchers neither report nor interpret their estimates of effect size. One noteworthy exception was Baggs and Brander's (2006: 207) attempt to convey in plain English the "economic significance" of the effects of trade liberalization in Canada:

> The effect of a large import tariff reduction reduces profit by \$146,000 for an average firm. At this rate, many firms protected by initially large tariffs would have profits reduced to zero over the phase-in period.

The beauty of statements such as these is in the way in which they convey information about the size of an observed effect that is meaningful to non-specialists, and uncomplicated by subsidiary issues of statistical significance. Regrettably, examples of this practice were hard to find.

RECOMMENDATIONS FOR RESEARCHERS

The widespread practice of interpreting p values as evidence in support of hypothesized effects constitutes a blatant disregard for the limitations of statistical significance testing. In other disciplines, such as psychology and education research, these limitations are better known, and have led to a number of reforms in research and reporting practices (Carver, 1978; Cortina & Dunlap, 1997). So far, researchers in the business disciplines have proven to be relatively slow adopters of these reforms, despite the exhortations of academy presidents and other opinion leaders.

In the hope of encouraging researchers to evaluate the substantive significance of their results, I offer four recommendations for improving research and reporting practices (Table 2). These

Table 2 Four recommendations for researchers

Recommendation	Recommending authority			
	JARS (2008)	*APA (2010)*	*AERA (2006)*	*Journal editors*
Analyze statistical power	✔ (p 842)	✔ (p 30)		Campbell (1982); Campion (1993); Combs (2010)
Report effect sizes	✔ (p 843)	✔ (p 34)	✔ (p 10)	Campion (1993); Combs (2010); Iacobucci (2005); JEP (2003); Shaver (2006); Zedeck (2003)
Report confidence intervals	✔ (p 843)	✔ (p 34)	✔ (p 10)	Campion (1993); Iacobucci (2005); Zedeck (2003)
Interpret substantive significance	✔ (implied)	✔ (p 35)	✔ (p 10)	Campbell (1982); Combs (2010); Rynes (2007); Shaver (2006)

recommendations are made with no agenda other than aiding the interpretation of quantitative results obtained from research samples. If international business researchers are able to compute the probability of detecting effects given their sample sizes (Recommendation 1), and if they then explicitly report both their obtained effect size estimates (Recommendation 2) and corresponding information regarding the precision of those estimates (Recommendation 3), they will be well placed to interpret the substantive significance of their results (Recommendation 4).

1. *Consider statistical power when designing studies*
Studies that have too much or too little statistical power are inherently wasteful, and potentially misleading. Even if researchers are careful to avoid making Type II errors, any underpowered study will lead to an inconclusive and therefore unsatisfactory result. Consequently the *Publication Manual* of the American Psychological Association recommends that authors "provide evidence that the study has sufficient power to detect effects of substantive interest" (APA, 2010: 30). Evidence in this context would be a prospective analysis of statistical power based on the anticipated effect size. Prior expectations regarding the effect size should be informed either by theory or by past research. Power analyses should not be based on effect sizes observed in the study, as study-specific estimates are rarely identical to the population effect size.

When researchers have little choice but to rely on relatively small samples, statistical power considerations should motivate them to seek out large effects. A good example of this comes from Van de Vliert's (2003) analysis of the link between mastery-oriented culture and wages. In this study the sample size was dictated by the relatively small number of countries (N=58) for which cultural data were available. With such a small sample the minimum effect size that could be detected using two-tailed tests with α set at 0.05 and power set at the bare minimum of 0.50 was r=0.25. Fortunately the effect observed in the study was larger than this (r=0.31), and the results achieved statistical significance.

2. *Report effect size estimates*
In the introduction to the fifth edition of its *Publication Manual* the APA identified the "failure to report effect sizes" as one of seven common defects that editors observed in submitted manuscripts (APA, 2001: 5). To help readers understand the importance of a study's findings, authors are now advised that "it is almost always necessary to include some measure of effect size in the Results Section" (APA, 2010: 34). Similarly, in its *Standards for Reporting*, the American Educational Research Association recommends that the reporting of statistical results should be accompanied by an "index of the quantitative relation between the variables" – that is, an estimate of the effect size (AERA, 2006: 10). These recommendations have been echoed by the editors of the *Journal of Applied Psychology* (Zedeck, 2003), the *Journal of Educational Psychology* (JEP, 2003), and *Personnel Psychology* (Campion, 1993). Among the business disciplines only a few journal editors have so far advocated the reporting of effect size estimates. This group includes Iacobucci (2005), writing for the *Journal of Consumer Research*, Shaver (2006), in a guest editorial for *JIBS*, and Combs (2010), recently writing in the *Academy of Management Journal*.

Effect size reporting is an essential precursor to meaningful interpretation (Ellis, 2010b). The reporting of effect sizes in widely understood metrics also promotes the cross-fertilization of ideas, an outcome that should be particularly desirable in an interdisciplinary journal such as *JIBS*.

3. *Quantify the precision of the estimate*
Along with a point estimate of the effect size, researchers should quantify the precision of their estimate by providing standard errors or confidence intervals. A confidence interval conveys more information than a p value, because it indicates the range of plausible values for the index being estimated. The wider the interval, the less confidence one should place in a point estimate of an effect size. Consequently the APA recommends the use of confidence intervals as "the best reporting strategy" (APA, 2010: 34). Examples of using standard errors or confidence intervals to quantify the uncertainty of an effect size in *JIBS* include Qian et al. (2008: Table 5).

4. *Interpret substantive significance*
The purpose of reporting estimates of the effect size along with their corresponding confidence intervals is so that authors might draw conclusions about real-world effects. Here the relevant question is not "How big is it?" or "How precise is the estimate?" but "What does it mean, and to whom?" In this regard the APA (2010: 35) exhorts authors to "evaluate and interpret" the implications of their results, while the AERA (2006: 10) calls for "a qualitative interpretation of the index of the effect".

Effect size indexes are meaningless unless they can be contextualized against some frame of reference. At a minimum, authors should interpret their results in the context of current evidence. Does the observed effect differ from what others have found and, if so, why and by how much? Writing in an editorial for the *Academy of Management Journal*, Rynes (2007: 1046) has also suggested that authors contextualize their results in terms of how they might "change or add to what we tell students in a classroom or managers in a consulting situation". A growing number of guidelines has recently emerged to help authors deal with the interpretation challenge (e.g., Cumming & Finch, 2005; Ellis, 2010a; Shaver, 2008).

CONCLUSIONS

Across the business disciplines there is an ongoing push towards relevance and engagement with stakeholders beyond the research community. Academy presidents and journal editors alike are calling for research that is "scientifically valid and practical" (Cummings, 2007: 355), and which culminates in the reporting of effect sizes that are "simultaneously helpful to academics, educators, and practitioners" (Rynes, 2007: 1048). These calls have largely gone unheeded. This is evident in the confusion of statistical with substantive significance, in the general neglect of effect sizes and statistical power, and in the corresponding lack of meaningful interpretation.

In the international business domain, average effects tend to be small to very small in size. However, most studies lack the power to detect such effects reliably. This makes the conduct of research something of a lottery and, when combined with an availability bias favoring statistically significant results, leads to the publication of non-replicable findings.

If we are to conduct research that matters, it is essential that we ensure our samples have the power to detect sought-after effects, that our effect size estimates are accurate, and that we don't shy away from making judgments about what those effects actually mean for people in the real world. Given the rising tide of reform in other disciplines, it is increasingly likely that contributions to *JIBS* will be gauged not merely in terms of whether they attain arbitrary standards of statistical significance, but in terms of how they affect our understanding of real-world effects.

ACKNOWLEDGEMENTS

I am grateful to Juergen Brock, Sylvie Chetty, Flora Gu, Evert Van de Vliert, Robert Wright, Consulting Editor Myles Shaver, and three anonymous *JIBS* referees for providing feedback and constructive criticism on earlier drafts of this paper. The research reported in this paper was supported by a Internal Research Grant provided by the Hong Kong Polytechnic University (Project 4-ZZ9V).

REFERENCES

Abraham, W. T., & Russell, D. W. 2008. Statistical power analysis in psychological research. *Social and Personality Compass*, 2(1): 283–301.

AERA. 2006. Standards for reporting on empirical social science research in AERA publications. http://www.aera.net/opportunities/?id=1850. Accessed 11 September 2008.

APA. 2001. *Publication manual of the American Psychological Association*, (5th ed.). Washington, DC: American Psychological Association.

APA. 2010. *Publication manual of the American Psychological Association*, (6th ed.). Washington, DC: American Psychological Association.

Baggs, J., & Brander, J. A. 2006. Trade liberalization, profitability, and financial leverage. *Journal of International Business Studies*, 37(2): 196–211.
Barden, J. Q., Steensma, H. K., & Lyles, M. A. 2005. The influence of parent control structure on parent conflict in Vietnamese international joint ventures: An organizational justice-based contingency approach. *Journal of International Business Studies*, 36(2): 156–174.
Brock, J. 2003. The "power" of international business research. *Journal of International Business Studies*, 34(1): 90–99.
Campbell, J. P. 1982. Editorial: Some remarks from the outgoing editor. *Journal of Applied Psychology*, 67(6): 691–700.
Campion, M. A. 1993. Article review checklist: A criterion checklist for reviewing research articles in applied psychology. *Personnel Psychology*, 46(3): 705–718.
Cano, C. R., Carrillat, F. A., & Jaramillo, F. 2004. A meta-analysis of the relationship between market orientation and business performance: Evidence from five continents. *International Journal of Research in Marketing*, 21(2): 179–200.
Carver, R. P. 1978. The case against statistical significance testing. *Harvard Educational Review*, 48(3): 378–399.
Cashen, L. H., & Geiger, S. W. 2004. Statistical power and the testing of null hypotheses: A review of contemporary management research and recommendations for future studies. *Organizational Research Methods*, 7(2): 151–167.
Child, J., Chung, L., & Davies, H. 2003. The performance of cross-border units in China: A test of natural selection, strategic choice and contingency theories. *Journal of International Business Studies*, 34(3): 242–254.
Cohen, J. 1988. *Statistical power for the behavioral analysis*, (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
Combs, J. G. 2010. Big samples and small effects: Let's not trade relevance and rigor for power. *Academy of Management Journal*, 53(1): 9–13.
Cortina, J. M., & Dunlap, W. P. 1997. On the logic and purpose of significance testing. *Psychological Methods*, 2(2): 161–172.
Cumming, G., & Finch, S. 2005. Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, 60(2): 170–180.
Cummings, T. G. 2007. 2006 presidential address: Quest for an engaged academy. *Academy of Management Review*, 32(2): 355–360.
Ellis, P. D. 2007. Distance, dependence and diversity of markets: Effects on market orientation. *Journal of International Business Studies*, 38(3): 374–386.
Ellis, P. D. 2008. Does psychic distance moderate the market size–entry sequence relationship? *Journal of International Business Studies*, 39(3): 351–369.
Ellis, P. D. 2010a. Effect size FAQs. http://www.effectsizefaq.com. Accessed 2 June 2010.
Ellis, P. D. 2010b. *The essential guide to effect sizes: An introduction to statistical power, meta-analysis and the interpretation of research results*. Cambridge: Cambridge University Press.
Hambrick, D. C. 1994. 1993 Presidential Address: What if the academy actually mattered? *Academy of Management Review*, 19(1): 11–16.
Hoenig, J. M., & Heisey, D. M. 2001. The abuse of power: The pervasive fallacy of power calculations for data analysis. *The American Statistician*, 55(1): 19–24.
Iacobucci, D. 2005. From the editor. *Journal of Consumer Research*, 32(1): 1–6.
JARS. 2008. Reporting standards for research in psychology. *American Psychologist*, 63(9): 839–851.
JEP. 2003. Instructions to authors. *Journal of Educational Psychology*, 95(1): 201.
Luk, C. L., Yau, O., Sin, L., Tse, A., Chow, R., & Lee, J. 2008. The effects of social capital and organizational innovativeness in different institutional contexts. *Journal of International Business Studies*, 39(4): 589–612.
Mazen, A. M., Graf, L. A., Kellogg, L. A., & Hemmasi, M. 1987. Statistical power in contemporary management research. *Academy of Management Journal*, 30(2): 369–380.
Parboteeah, K. P., Hoegl, M., & Cullen, J. B. 2008. Managers' gender role attitudes: A country institutional profile approach. *Journal of International Business Studies*, 39(5): 795–813.
Qian, G. M., Li, L., Li, J., & Qian, Z. M. 2008. Regional diversification and firm performance. *Journal of International Business Studies*, 39(2): 197–214.
Rynes, S. L. 2007. Editors afterword: Let's create a tipping point – What academics and practitioners can do, alone and together. *Academy of Management Journal*, 50(5): 1046–1054.
Scherbaum, C. A., & Ferreter, J. M. 2009. Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods*, 12(2): 347–367.
Shaver, J. M. 2006. Interpreting empirical findings. *Journal of International Business Studies*, 37(4): 451–452.
Shaver, J. M. 2008. Organizational significance. *Strategic Organization*, 6(2): 185–193.
Van de Vliert, E. V. 2003. Thermoclimate, culture, and poverty as country-level roots of workers' wages. *Journal of International Business Studies*, 34(1): 40–52.
Zedeck, S. 2003. Editorial. *Journal of Applied Psychology*, 88(1): 3–5.

ABOUT THE AUTHOR

Paul D Ellis is a Professor at the Hong Kong Polytechnic University. Born in New Zealand, he is a dual citizen of New Zealand and Australia, and a permanent resident of Hong Kong. He received his PhD from the University of Western Australia. His research interests include trade intermediation, marketing and economic development, the geography of international business, and statistical power issues. E-mail: mspaul@inet.polyu.edu.hk.

Accepted by Myles Shaver, Consulting Editor, 25 July 2010. This paper has been with the author for three revisions.