

Research in the Psychological Laboratory: Truth or Triviality?

Craig A. Anderson,¹ James J. Lindsay, and Brad J. Bushman

Department of Psychology, University of Missouri—Columbia, Columbia, Missouri (C.A.A., J.J.L.), and Department of Psychology, Iowa State University, Ames, Iowa (B.J.B.)

Abstract

This article examines the truism that studies from psychological laboratories are low in external validity. Past rational and empirical explorations of this truism found little support for it. A broader empirical approach was taken for the study reported here; correspondence between lab and field was compared across a broad range of domains, including aggression, helping, leadership style, social loafing, self-efficacy, depression, and memory, among others. Correspondence between lab- and field-based effect sizes of conceptually similar independent and dependent variables was considerable. In brief, the psychological laboratory has generally produced psychological truths, rather than trivialities. These same data suggest that a companion truism about field studies in psychology—that they are generally low on internal validity—is also false.

Keywords

external validity; metaanalysis; philosophy of science

Are you happy with current rates of violent crime in the United States? How about the U.S. ranking in achievement test scores in science, compared with other industrialized nations? Do recent increases in smoking rates among U.S. teens bother you? Do the continuing problems of high rates of

“unsafe sex” practices and the resulting incidence of AIDS seem to cry out for a solution?

Constant attention to the problems of modern U.S. society by the mass media, politicians, and concerned citizens may seem to indicate a generally pessimistic worldview. Paradoxically, though, this focus on problems actually reflects a fundamentally optimistic view that as a society we can and should solve these problems. This same optimism drives modern psychology as well. The whole point of the science of psychology is to learn so that we can improve “things.”

THE DEBATE ABOUT EXTERNAL VALIDITY

This functionalist view gives rise to a long-running and frequently counterproductive debate about the value of theory-oriented laboratory research versus application-oriented field research. This article addresses a very specific question from this debate: Does the psychological laboratory yield truths or trivialities? Since Campbell (1957) clearly distinguished between *internal* and *external validity*, a common truism has been that laboratory studies are typically high on internal validity but low on external validity.² That is, laboratory studies are good at telling whether or not some manipulation of an independent variable causes changes in the dependent variable, but many scholars assume that these results do not generalize to the “real world.” Hence, application-

oriented scholars sometimes deride the psychological laboratory as a place where only trivial facts are to be found.³ In essence, the charge is that (some, most, or all) research from the psychological laboratory is externally invalid, and therefore pointless.

One domain where this debate periodically arises concerns aggression (Anderson & Bushman, 1997). Consider obvious differences in surface characteristics between real-world versus laboratory aggression. Assault typically involves two or more people who know each other, arises from an escalating cycle of provocation, and results in serious physical injury. Aggression in the lab, however, typically involves one person (who only thinks that he or she is interacting with an unknown person via computer, or notes, or message boards of one kind or another), in a session that may last only 50 min, and involves the attempted delivery of noxious stimuli such as electric shock or blasts of noise.

But the charge that psychological laboratory research lacks external validity is not unique to the study of aggression. Recent years have seen similar debates in the study of personnel selection (e.g., Schmitt, 1996), leadership (e.g., Wolfe & Roberts, 1993), management (e.g., Griffin & Kacmar, 1991), and human memory (e.g., Banaji & Crowder, 1989; Neisser, 1978). For instance, in one early laboratory study of context-dependent memory, Dallett and Wilcox (1968) asked participants in one condition to study word lists while standing with their heads inside an oddly shaped box that contained flashing lights of several different colors; other participants studied the word lists without putting their heads inside the box. Participants later recalled the lists either with or without the box, and were most successful when the study and recall conditions were the same. As is obvious, these condi-

tions have no counterpart in the "real" world, thus inviting complaints about external validity.

It is easy to see why nonexperts frequently charge that lab studies are trivial, artificial, and pointless, and easy to ignore such complaints as reflections of ignorance. But when the charge comes from experts—other psychological researchers who presumably share goals, training, and perspective—a thoughtful response is required. Such responses have also been forthcoming.

RESPONSES TO LABORATORY CRITICISMS

Embracing Invalidity

One particularly elegant response, by Mook (1983), celebrates external *invalidity*. He described four cases in which the artificial lab setting is not only acceptable but actually preferred to the real-world setting:

First, we may be asking whether something can happen, rather than whether it typically does happen. Second, our prediction may . . . specify something that ought to happen in the lab. . . . Third, we may demonstrate the power of a phenomenon by showing that it happens even under unnatural conditions that ought to preclude it. Finally, we may use the lab to produce conditions that have no counterpart in real life at all . . . (p. 382)

Mook's main point is that the goal of most laboratory research is to discover theoretical relations among conceptual variables that are never sufficiently isolated in the real world to allow precise examination.

What Is Supposed to Generalize?

A second (and related) response is to note that usually researchers are interested in generalization of theoretical relations among concep-

tual independent and dependent variables, not the specific instantiations of them. The same scandalous joke will mean something different in church than it does in the men's locker room. In one case it may create embarrassment, whereas in the other it may create humor. If one were interested in the effects of humor on thought processes, one would be foolish to use the same joke as an experimental manipulation of "humor" in both settings. The lack of a manipulation's generalizability constitutes an external validity problem only if one intends specific instantiations of conceptual variables to generalize across radically different contexts, but most laboratory research is concerned only with generalizability of the conceptual variables.

The General Problem With Generalization

A third response begins with the observation that generalization is, generally, risky business. Inductive reasoning, at least in absolute terms, is never wholly justified. Even though every time you have dropped a hammer it has fallen, you cannot know for certain that it will fall the next time. Perhaps the laws of nature will change, or perhaps you will enter a location where your understanding of the laws of nature is revealed to be incomplete. Thus, generalizing from one situation to another, or from one participant population to another, is as problematic for field research as it is for lab research. So, the argument goes, why single out lab research for criticism? At least the lab makes it somewhat easier to satisfy concerns about internal validity, and without internal validity there is nothing to generalize anyway.

But, people do generalize from specific instances to general concepts, then from these general concepts to new situations involving different instances of the general

concepts. A justification for such generalization can be readily found both at the level of species survival and at the level of scientific and technological advances: In both cases, generalization works much of the time.

Past Empirical Approaches

If what psychologists expect (hope?) to generalize are systematic relations among conceptual variables (i.e., theories), and if we grant that attempting to make generalizations is acceptable as long as it occasionally works, then another response to the external validity challenge becomes feasible. The challenge becomes an empirical question. Three different empirical approaches have been used: single-study tests, single-phenomenon tests, and single-domain tests.

Single-study tests examine a specific laboratory finding in other contexts or with other populations. For example, Godden and Baddeley (1975) successfully generalized the context-dependent memory effect using scuba divers as subjects. Word lists that had been studied underwater were better recalled underwater, whereas lists studied on dry land were better recalled on dry land. Such single-study tests of external validity abound in psychology. Many "work," though of course some do not. Though these tests answer the generalization question for a particular case, they do not adequately answer the broader question concerning the external validity of a given laboratory phenomenon.

Single-phenomenon tests examine the external validity of a whole empirical phenomenon rather than one specific laboratory finding. For example, do laboratory-based effects of anonymity on aggression generalize to field settings? Questions like this can be investigated by using *meta-analytic* techniques. That is, one could statisti-

cally average all of the research results from lab and field studies that have tested the relation between anonymity and aggression. We performed such a meta-analysis (Anderson & Bushman, 1997) and found comparable anonymity effects in lab and field settings. Similar tests of the generalizability of a specific laboratory phenomenon can be found in numerous additional areas of psychology. Many of these single-phenomenon tests show comparable effects for the psychological laboratory and field studies. But failures to generalize also occur.

Single-domain tests further broaden the generalizability question to a whole research domain. For example, do most aggression findings from the psychological laboratory generalize to field studies? In other words, do the effects of key independent variables—such as alcohol, anonymity, and media violence—have the same effects in lab and field studies of aggression? If laboratory studies from a given domain are inherently invalid and those from field studies are valid, then lab and field studies in that domain should fail to show any correspondence. We (Anderson & Bushman, 1997; Bushman & Anderson, 1998) used meta-analytic techniques to ask this broad question in the aggression domain, and found considerable correspondence between lab and field.

A CROSS-DOMAIN EMPIRICAL APPROACH

We extend the single-domain approach by examining the comparability of findings from lab and field across several domains. Basically, we asked whether the effects of the same conceptual independent variables on the same conceptual dependent variables tended to be consistent in lab and

field settings across several psychological domains.

Method

Using the PsycINFO database, we conducted a literature search for the following journals: *Psychological Bulletin*, *Journal of Applied Psychology*, *Journal of Personality and Social Psychology*, and *Personality and Social Psychology Bulletin*. We searched with the keyword phrases “meta-analysis” and “quantitative review,” and with the combined keyword phrases “meta-analysis” with “field studies” and “meta-analysis” with “laboratory.” This selection of journals was intentionally biased toward social psychology because many of the most vociferous criticisms of the psychological laboratory have focused on the social psychological lab. Our search yielded 288 articles.

Many articles were subsequently eliminated because they were methodological in nature, did not include separate tabulations for lab and field settings, or overlapped with a more recent meta-analytic review.⁴ The final data set represents 38 pairs of lab and field effects.

Results and Discussion

We used the standardized mean difference, denoted by d , as the indicator of effect size. This index shows the size of the difference between two groups, and does so in terms of the standard deviation. For instance, if Group 1 has a mean of 6 and Group 2 has a mean of 5, and the standard deviation is 2, the effect size d would be $(6 - 5)/2 = 0.5$. According to Cohen (1988), a “large” d is 0.8, a “medium” d is 0.5, and a “small” d is 0.2. Effect sizes for correlations can easily be converted to d s, which we did to allow direct comparisons across different types of studies (Hedges & Olkin, 1985). Effect-size averages were weighted by sample size and used

pooled standard deviations in most cases, but in a few cases we were unable to determine the weights from the original reports. Table 1 and Figure 1 summarize the results.⁵

Figure 1 plots the value of d for the lab and field studies for each domain studied. The figure reveals considerable consistency between laboratory and field effects. That is, across domains, the d s for lab and field studies tended to be similar. The correlation, $r = .73$, is considerably higher than the gloomy picture that sometimes emerges from the external validity debate. Some readers might wonder whether the disproportionate number of the data points coming from comparisons of gender effects (6 out of 38) biased the results. However, the plot and correlation look much the same with these 6 data points eliminated ($r = .75$). Similarly, one might wonder about the possible nonindependence of some of the attributional-style data points (the results from Sweeney, Anderson, & Bailey, 1986). Dropping all but the two overall “attribution and depression” effects for positive and negative outcomes again yields essentially the same correlation ($r = .73$). All three of these correlations are considerably larger than Cohen’s (1988) conventional value for a large correlation ($r = .5$). Furthermore, an r of .73 is equivalent to a d of 2.14, a huge effect.

Two complementary questions arise from these results, one asking why the correspondence is so high, the other asking why it is not higher. First, consider some limitations on correspondence between lab and field results. One limitation concerns internal validity problems of field studies. Sometimes field studies “discover” relations between independent and dependent variables that are false, and at other times they fail to discover true relations. Both types of internal validity reduce correspondence

Table 1. Mean effect sizes and confidence intervals for topics studied in the lab and field

Source, independent and dependent variables, and setting	Number of samples	Effect size	95% confidence interval
Ambady and Rosenthal (1992)			
Observation time and outcome ratings—lab	21	0.87	—
Observation time and outcome ratings—field	17	0.98	—
Anderson and Bushman (1997)			
Gender and physical aggression—lab	37	0.31	0.23–0.38
Gender and physical aggression—field	6	0.40	0.25–0.55
Gender and verbal aggression—lab	18	0.13	0.03–0.24
Gender and verbal aggression—field	3	0.03	–0.15–0.22
Bushman and Anderson (1998)			
Anonymity and aggression—lab	18	0.57	0.45–0.69
Anonymity and aggression—field	4	0.44	0.25–0.63
Trait aggressiveness and aggression—lab	13	0.49	0.18–0.29
Trait aggressiveness and aggression—field	16	0.93	0.38–0.47
Type A personality and aggression—lab	9	0.34	0.18–0.49
Type A personality and aggression—field	3	0.97	0.71–1.23
Carlson, Marcus-Newhall, and Miller (1990)			
Weapons and aggression—lab	16	0.21	0.01–0.41
Weapons and aggression—field	5	0.17	–0.05–0.39
Eagly and Crowley (1986)			
Gender and helping—lab	16	–0.18	–0.28–0.09
Gender and helping—field (on and off campus)	36 + 47 = 83	0.27	—
Eagly and Johnson (1990)			
Gender and leadership style—lab	17	0.22	—
Gender and leadership style—organizations	269	–0.00	—
Eagly and Karau (1991)			
Gender and leader emergence—lab	50	0.45	0.40–0.51
Gender and leader emergence—natural settings	24	0.10	0.02–0.17
Eagly, Karau, and Makhijani (1995)			
Gender and leader effectiveness—lab	20	0.07	–0.06–0.20
Gender and leader effectiveness—organizations	56	–0.03	–0.06–0.01
Gordon (1996)			
Ingratiation and evaluations—university (lab)	54	0.38	0.33–0.43
Ingratiation and evaluations—field	15	–0.07	–0.13–0.00
Karau and Williams (1993)			
Social loafing—lab	140	0.47	0.43–0.51
Social loafing—field	23	0.25	0.16–0.35
Kraiger and Ford (1985)			
Race of ratee and performance ratings—lab	10	0.07	–0.41–0.56
Race of ratee and performance ratings—field	64	0.39	0.06–0.75
Kubeck, Delp, Haslett, and McDaniel (1996)			
Age (continuous) and job-training mastery—lab	17	–0.61	–1.67–0.12
Age (continuous) and job-training mastery—field	31	–0.52	–1.19–0.02
Age and time to finish training—lab	3	0.70	0.08–1.58
Age and time to finish training—field	2	1.35	1.35–1.35
Age (dichotomous) and job-training mastery—lab	9	–0.96	–1.44–0.47
Age (dichotomous) and job-training mastery—field	2	–0.38	–0.38–0.38
Lundeberg and Fox (1991)			
Expectancies and recall—essay tests—lab	41	0.60	0.53–0.67
Expectancies and recall—essay tests—class	11	0.33	0.17–0.49
Expectancies and recognition tests—lab	41	–0.07	–0.13–0.01
Expectancies and recognition tests—class	14	0.28	0.14–0.42
Mento, Steel, and Karren (1987)			
Goal difficulty and performance—lab	47	0.62	—
Goal difficulty and performance—field	23	0.44	—
Mullen and Hu (1989)			
Group membership and similarity of group members— artificially created groups	2	0.43	0.04–0.90
Group membership and similarity of group members— real groups	2	0.47	–0.14–1.16

Table 1. *continued*

Source, independent and dependent variables, and setting	Number of samples	Effect size	95% confidence interval
Narby, Cutler, and Moran (1993)			
Authoritarianism and trial verdict—video, written, audio trials	23	0.30	—
Authoritarianism and trial verdict—live trials	3	0.49	—
Paik and Comstock (1994)			
Media violence and aggression—lab	586	0.87	—
Media violence and aggression—field	556	0.42	—
Peters, Hartke, and Pohlman (1985)			
Leadership style and performance, negative octants—lab	30	-0.28	—
Leadership style and performance, negative octants—field	20	-0.90	—
Leadership style and performance, positive octants—lab	20	0.51	—
Leadership style and performance, positive octants—field	15	0.45	—
Sagie (1994)			
Decision-making participation and productivity—lab	—	-0.06	—
Decision-making participation and productivity—field	—	-0.02	—
Sweeney, Anderson, and Bailey (1986)			
Attribution and depression, negative outcomes—lab	25	0.52	—
Attribution and depression, negative outcomes—hospital	8	0.32	—
Attribution and depression, positive outcomes—lab	16	-0.24	—
Attribution and depression, positive outcomes—hospital	5	-0.28	—
Ability and depression, negative outcomes—lab	16	0.63	—
Ability and depression, negative outcomes—hospital	3	1.15	—
Ability and depression, positive outcomes—lab	13	-0.24	—
Ability and depression, positive outcomes—hospital	3	-0.12	—
Effort and depression, negative outcomes—lab	13	0.10	—
Effort and depression, negative outcomes—hospital	2	0.49	—
Effort and depression, positive outcomes—lab	11	-0.02	—
Effort and depression, positive outcomes—hospital	2	-0.04	—
Luck and depression, negative outcomes—lab	14	-0.30	—
Luck and depression, negative outcomes—hospital	3	-0.61	—
Luck and depression, positive outcomes—lab	10	0.43	—
Luck and depression, positive outcomes—hospital	3	0.63	—
Task difficulty and depression, negative outcomes—lab	14	-0.26	—
Task difficulty and depression, negative outcomes—hospital	2	-0.14	—
Task difficulty and depression, positive outcomes—lab	9	-0.20	—
Task difficulty and depression, positive outcomes—hospital	2	0.61	—
Tubbs (1986)			
Goal specificity and performance—lab	34	0.57	0.14–1.01
Goal specificity and performance—field	14	0.43	-0.09–0.94
Goal-setting participation and performance—lab	13	-0.03	-0.86–0.80
Goal-setting participation and performance—field	4	0.12	-0.34–0.59

Note. All effect-size estimates have been converted to d —the average effect size in standard deviation units, weighted by sample size whenever sufficient information was available to do so. If the exact same sampling and methodological procedures were used to gather new data to estimate d , and if this were done a large number of times, we should expect that 95% of the time, the new d estimates would fall within the range indicated by the 95% confidence interval.

between field and lab, and hence artificially depress the correlation seen in Figure 1. A second limitation concerns the primary reason for studying psychological phenomena in the lab—to improve one's ability to detect relatively subtle phenomena that are difficult or impossible to isolate in the field. Laboratory studies typically accomplish this by focusing on one or two theoretically interesting inde-

pendent variables while restricting the action or range of other independent variables. This focus can increase the estimated effect size of experimentally manipulated independent variables while decreasing the effects of individual difference variables. For example, using only college students as experimental participants restricts the range of individual differences in intelligence and antisocial personality,

thereby reducing their effects on dependent variables associated with them. Therefore, reducing the effects of individual differences in the lab also reduces variance due to chance or measurement error and thus increases the estimated effect size of manipulated variables, relative to the estimated effect sizes generated from similar field studies without the (intentional) range restriction on subject variables.

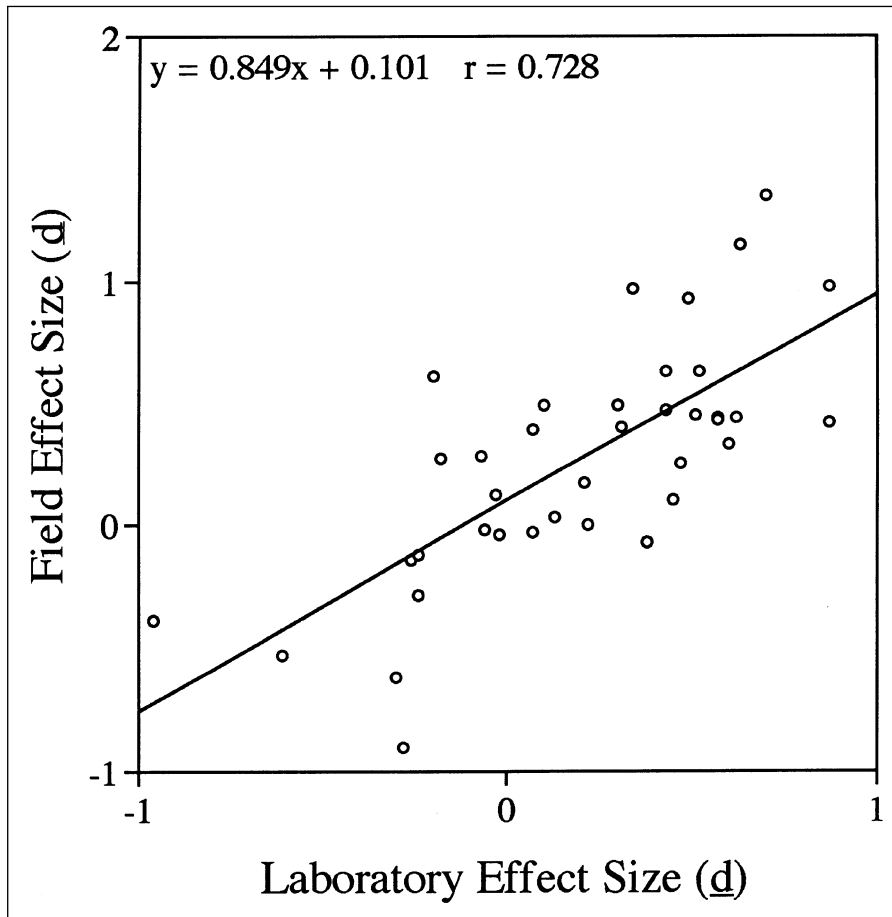


Fig. 1. Relation between effect sizes in the laboratory and field. Each point represents the value of d for the lab and field studies in a particular meta-analysis.

Thus, both internal validity problems of field studies and the range restriction of lab studies artificially decrease the lab-field correspondence displayed in Figure 1.

Now consider the other question, about factors that helped make the correspondence in Figure 1 so high. First, meta-analytically derived indicators of effect size wash out idiosyncratic effects of individual studies. That is, random or idiosyncratic factors that artificially increase an effect in some studies and decrease it in others tend to balance out when averaged across many studies, in much the way that increasing sample size increases the accuracy of the results in a single study. Second, and perhaps more important, we investigated only research domains that have had sufficient research attention to allow meta-analyses. These

would usually be successful research domains, where underlying theories and methods are accurate enough to produce a line of successful studies. Such successful lines of investigation are relatively likely to concern true (internally valid) relations between the key independent and dependent variables.

CONCLUSIONS

The obvious conclusion from Figure 1 is that the psychological laboratory is doing quite well in terms of external validity; it has been discovering truth, not triviality. Otherwise, correspondence between field and lab effects would be close to zero.

A less obvious conclusion concerns internal validity of field stud-

ies. A second part of the broader debate between theory-oriented laboratory researchers and application-oriented field researchers is the truism that field studies generally lack internal validity. If this second truism were accurate, however, the correspondence between lab and field effects could not have been so positive. Thus, field studies in psychology must be doing a pretty good job when it comes to internal validity.

In the interests of clarity and space, we have oversimplified the lab-field debate on validity. Obviously, studies in either setting may be high or low on internal and external validity. As long as scholars in both settings keep in mind the complementary pitfalls of too little control over extraneous variables (leading to low internal validity) and of overgeneralizing from the specific features of a specific study (leading to low external validity), we believe the psychological research enterprise will continue to succeed.

Finally, failure to find high correspondence between lab and field studies in a given domain or with a specific phenomenon should not be seen as a failure of the researchers in either setting. Instead, such inconsistencies should be seen as an indicator that further conceptual analysis and additional empirical tests are needed to discover the source of the discrepancy. Perhaps there are psychological processes operating in one context but not the other, or perhaps the relative strength of different causal factors differs in the two contexts (see Anderson & Anderson, 1998, for an example involving the positive relation between uncomfortably hot temperatures and aggressive behavior). In any case, the discrepancy sets the stage for further theoretical and (eventually) practical advances. And in the end, that's what we all are working for, isn't it?

Recommended Reading

- Anderson, C.A., & Bushman, B.J. (1997). (See References)
- Banaji, M.R., & Crowder, R.G. (1989). (See References)
- Berkowitz, L., & Donnerstein, E. (1982). External validity is more than skin deep: Some answers to criticism of laboratory experiments. *American Psychologist*, *37*, 245–257.
- Kruglanski, A.W. (1975). The human subject in the psychology experiment: Fact and artifact. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8, pp. 101–147). New York: Academic Press.
- Mook, D.G. (1983). (See References)

Acknowledgments—We thank Bruce Bartholow and Anne Powers for comments on an earlier version. This article is based in part on an invited address delivered by the first author at the May 1998 meeting of the American Psychological Society in Washington, D.C.

Notes

1. Address correspondence to Craig A. Anderson, Department of Psychology, Iowa State University, W112 Lagomarcino Hall, Ames, IA 50011-3180.

2. Internal validity refers to the degree to which the design, methods, and procedures of a study allow one to conclude that the independent variable caused observable differences in the dependent variable. External validity refers to the degree to which the relationship between the independent and dependent variables found in a study generalizes to other people, places, and times.

3. The companion truism, held by some scholars with a more theoretical orientation, is that field studies on “real” phenomena are so plagued by methodological confounds that they lack internal validity, and, hence, fail to say anything at all about the phenomenon under study.

4. For example, the Wood, Wong, and Cachere (1991) analysis of the effects of violent media on aggression

overlaps with Paik and Comstock’s (1994) analysis.

5. Two additional meta-analyses also demonstrated considerable lab-field correspondence, but did not report separate effect sizes. Kraus’s (1995) review of the consistency between attitude and behavior and Kluger and DeNisi’s (1996) review of the effects of feedback on performance both coded effects by the lab-field distinction, and both reported a nonsignificant relationship between this distinction and effect size.

References

- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*, 256–274.
- Anderson, C.A., & Anderson, K.B. (1998). Temperature and aggression: Paradox, controversy, and a (fairly) clear picture. In R. Geen & E. Donnerstein (Eds.), *Human aggression: Theories, research, and implications for social policy* (pp. 247–298). San Diego: Academic Press.
- Anderson, C.A., & Bushman, B.J. (1997). External validity of “trivial” experiments: The case of laboratory aggression. *Review of General Psychology*, *1*, 19–41.
- Banaji, M.R., & Crowder, R.G. (1989). The bankruptcy of everyday memory. *American Psychologist*, *44*, 1185–1193.
- Bushman, B.J., & Anderson, C.A. (1998). Methodology in the study of aggression: Integrating experimental and nonexperimental findings. In R. Geen & E. Donnerstein (Eds.), *Human aggression: Theories, research, and implications for social policy* (pp. 23–48). San Diego: Academic Press.
- Campbell, D.T. (1957). Factors relevant to validity of experiments in social settings. *Psychological Bulletin*, *54*, 297–312.
- Carlson, M., Marcus-Newhall, A., & Miller, N. (1990). Effects of situational aggression cues: A quantitative review. *Journal of Personality and Social Psychology*, *58*, 622–633.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Dallett, K., & Wilcox, S.G. (1968). Contextual stimuli and proactive inhibition. *Journal of Experimental Psychology*, *78*, 475–480.
- Eagly, A.H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin*, *100*, 283–285.
- Eagly, A.H., & Johnson, B.T. (1990). Gender and leadership style: A meta-analysis. *Psychological Bulletin*, *108*, 233–256.
- Eagly, A.H., & Karau, S.J. (1991). Gender and emergence of leaders: A meta-analysis. *Journal of Personality and Social Psychology*, *60*, 685–710.
- Eagly, A.H., Karau, S.J., & Makhijani, M.G. (1995). Gender and the effectiveness of leaders: A meta-analysis. *Psychological Bulletin*, *117*, 125–145.
- Godden, D., & Baddeley, A. (1975). When does context influence recognition memory? *British Journal of Psychology*, *71*, 99–104.
- Gordon, R.A. (1996). Impact of ingratiation on judgments and evaluations: A meta-analytic investigation. *Journal of Personality and Social Psychology*, *71*, 54–70.
- Griffin, R., & Kacmar, M.K. (1991). Laboratory research in management: Misconceptions and missed opportunities. *Journal of Organizational Behavior*, *12*, 301–311.
- Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analyses*. New York: Academic Press.
- Karau, S.J., & Williams, K.D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, *65*, 681–706.
- Kluger, A.N., & DeNisi, A. (1996). Effects of feedback on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254–284.
- Kraiger, K., & Ford, J.K. (1985). A meta-analysis of rater effects in performance ratings. *Journal of Applied Psychology*, *70*, 56–75.
- Kraus, S.J. (1995). Attitudes and the prediction of behavior: A meta-analysis of the empirical evidence. *Personality and Social Psychology Bulletin*, *21*, 58–75.
- Kubeck, J.E., Delp, N.D., Haslett, T.K., & McDaniel, M.A. (1996). Does job-related training performance decline with age? *Psychology and Aging*, *11*, 92–107.
- Lundeberg, M.A., & Fox, P.W. (1991). Do laboratory findings on test expectancy generalize to classroom outcomes? *Review of Educational Research*, *61*, 94–106.
- Mento, A.J., Steel, R.P., & Karren, R.J. (1987). A meta-analytic study of the effects of goal setting on task performance: 1966–1984. *Organizational Behavior & Human Decision Processes*, *39*, 52–83.
- Mook, D.G. (1983). In defense of external invalidity. *American Psychologist*, *38*, 379–387.
- Mullen, B., & Hu, L. (1989). Perceptions of ingroup and outgroup variability: A meta-analytic integration. *Basic and Applied Social Psychology*, *10*, 233–252.
- Narby, D.J., Cutler, B.L., & Moran, G. (1993). A meta-analysis of the association between authoritarianism and jurors’ perceptions of defendant culpability. *Journal of Applied Psychology*, *78*, 34–42.
- Neisser, U. (1978). Memory: What are the important questions? In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.), *Practical aspects of memory* (pp. 3–24). London: Academic Press.
- Paik, H., & Comstock, G. (1994). The effects of television violence on antisocial behavior: A meta-analysis. *Communication Research*, *21*, 516–546.
- Peters, L.H., Hartke, D.D., & Pohlman, J.T. (1985). Fiedler’s contingency theory of leadership: An application of the meta-analysis procedures of Schmidt & Hunter. *Psychological Bulletin*, *97*, 274–285.
- Sagie, A. (1994). Participative decision making and performance: A moderator analysis. *Journal of Applied Behavioral Science*, *30*, 227–246.
- Schmitt, N. (1996). Validity generalization. In R. Barrett (Ed.), *Fair employment strategies in human resource management* (pp. 94–104). Westport, CT: Quorum Books.
- Sweeney, P.D., Anderson, K., & Bailey, S. (1986). Attributional style in depression: A meta-analytic review. *Journal of Personality and Social Psychology*, *50*, 974–991.
- Tubbs, M.E. (1986). Goal setting: A meta-analytic examination of the empirical evidence. *Journal of Applied Psychology*, *71*, 474–483.
- Wolfe, J., & Roberts, C.R. (1993). A further study of the external validity of business games: Five-year peer group indicators. *Simulation & Gaming*, *24*, 21–33.
- Wood, W., Wong, F.Y., & Cachere, J.G. (1991). Effects of media violence on viewers’ aggression in unconstrained social interaction. *Psychological Bulletin*, *109*, 371–383.