The Scientific Basis of the Lack of replicability in Psychological Research

Thomas J. Smith (UM College of Education & Human Development School of Kinesiology)

The issue of lack of replicability of psychological research recently has attracted intense scrutiny. This presentation argues that the pervasive role of *context (e.g., design) specificity* in behavioral performance and learning poses a formidable human factors challenge to rigorous replicability in psychological research. Because of this role, it is hard to re-create the exact conditions of the original research. That is, many variables and conclusions in psychology cannot be fully understood apart from the design contexts that define their meanings and implications.

Key Conclusions: Experimental findings reviewed in this presentation support the following conclusions: (1) results of early research indicate that cognitive performance is task specialized in relation to transfer effects between different tasks, as well as in relation to the predictive significance of initial performance in forecasting fina1 performance; (2) results of subsequent research confirm and extend these conclusions by demonstrating that for most tasks, variability observed in cognitive performance is more prominently linked to task-specific factors than to general attributes of intelligence or ability; and (3) these findings pose a formidable scientific challenge to the replicability of psychological research, unless experimental analysis is conducted and/or validated in real world contexts.

1 Background - 1

- The science of psychology is facing a crisis [1]
- Many studies in psychology—including many highly cited studies—do not replicate
- The failure to replicate research raises important questions about the scientific process in general and psychology specifically
- People have the right to know if they can trust research
- Psychologists also have a vested interest in ensuring that their methods and findings are as trustworthy as possible.

2 Extent of Non-Replicability of Psychology Findings - Percentages of Research that Replicated - Selected from Several Highly Prestigious Journals [2; 270 authors]

Journal	% Findings Replicated
Journal of Personality and Social Psychology: Social	23
J. of Experimental Psychology: Learning, Memory & Cognition	48
Psychological Science, social articles	29
Psychological Science, cognitive articles	53
Overall	36

3 ► Background - 2

- Modern scientific insight into the challenge of replicating behavioral research observations dates back well over a century, to the studies of Thorndike & Woodworth [6-8]
- Key conclusions from this work are:

"Improvement in any single mental function need not improve the ability in functions commonly called by the same name. It may injure it."

"Improvement in any single mental function rarely brings about equal improvement in any other function, no matter how similar, for the working of every mental function-group is conditioned by the nature of the data in each particular case."

- In other words, within- or between-subject replicability of simple behavioral tasks (i.e., functions --- math, verbal, visual, etc.) is unpredictable and inconsistent

4 Possible Reasons for Non-Replicability of Psychological Research

- Over 70 years ago, Lewin [9] noted that behavior is influenced by both the person (P) and the environment (E): B = f(P,E)
- References 1-5, plus an entire 2012 issue of *Perspectives on Psychological Science* (Volume 7, Number 6), offer different perspectives on this question.
- ► Ioannidis [4, p. 650] lists possible reasons for lack of replicability, summarized below

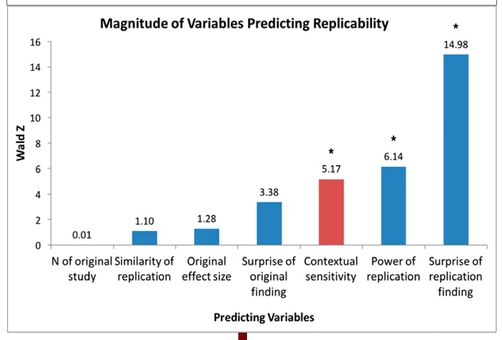
A List of Impediments to replicability in Science, with Reference to Psychological Science

Publication bias
Other selective reporting bias (analysis and outcomes)
Flexibility in data collection and analysis
Misreporting of results
Voodoo correlations
Fabricated results
Other questionable research practices
Excess significance bias (may reflect any of the above)
Underpowered studies
No replication work done—especially direct replication by independent investigators
Underestimation of the replication crisis
Editorial bias against replication research
Reviewer bias against replication research
Data, analyses, protocols not publicly available

5 Why should replicability in psychological research be compromised by design specificity in behavior and performance?

Recent Perspectives on This Question:

- Findings may only be true for some people (sampling design), and/or in some circumstances (study context) [1]
- Van Bevel et al. [5] analyzed 100 behavioral studies to develop an index called *contextual sensitivity,* based on the degree to which the following study design factors may have influenced replicability: 1) time (month or year); 2) culture; (3) location; or 4) sample population
- Contextual sensitivity was negatively correlated with the success of the replication attempt, ($r(98) = -0.23$, $p = 0.024$), such that the more contextually sensitive topic was rated, the less likely was the replication attempt to be successful
- Only two other variables (below) also significantly predicted replicability (positive or negative correlation)



6 Other contextual factors not addressed by Van Bevel et al. [5] that may also influence replicability

- Race or gender of experimenters or subjects
- Temperature
- Time of day

7 EVIDENCE FOR CONTEXT SPECIFICITY IN BEHAVIOR AND PERFORMANCE

SUMMARY FINDINGS --- AND CONCLUSIONS ---FROM SELECTED STUDIES

8 Poffenburger [9] --- Poor Transfer of Learning for Tasks with Highly Similar Designs

TABLE 1. Transfer of Learning in Cognitive Tasks*.

	Transfer Experiment			Type of Transfer
1.	Color Naming	--->	Form Naming	None
2.	Opposites Test	--->	Adjective Noun Test	Negative
3.	Two Digit Cancellation	--->	Group Cancellation	Positive
4.	Two Digit Cancellation	--->	Different Two Digits	None
5.	Addition	--->	Subtraction	None
6.	Addition	--->	Multiplication	Negative
7.	Addition	--->	Division	None

*Results of Poffenburger (1915)

Conclusions

Poor replicabiity in behavioral performance, even for tasks that appear to be very similar in design

EVIDENCE FOR CONTEXT SPECIFICITY IN BEHAVIOR AND PERFORMANCE (continued)

9 Dewey (1916) [10] --- Significance of Informal Learning

"From the standpoint of the child, the great waste in school comes from his inability to utilize the experience he gets outside…while on the other hand, he is unable to apply in daily life what he is learning in school. That is the isolation of the school—its isolation from life."

Conclusions

Context outside the classroom has more influence on student behavior and learning than classroom learning

10 Woodrow's (1939) [11] Concept of Specific Variance

Specific Variance refers to the influence of task structure or makeup (versus individual differences or learning) on cognitive performance

Conclusions

There is no general mental factor or process which accounts for improvement in cognitive learning.

The specific variance of the practiced tasks increases with practice, and this increase is a function of the nature (design) of the task

11 Extension of Woodrow's Work by the Task Taxonomy Research of Fleishman (1966) [12]

Studies of a great variety of practice tasks show that:

- The particular combination of abilities contributing to performance changes as practice continues;
- These changes are progressive and systematic and eventually become stabilized;
- As practice continues, *there is an increase in a factor specific to the task itself* (termed task-specific variance)

12 Refinement of Fleishman's Work by Ackerman (1987) [13]

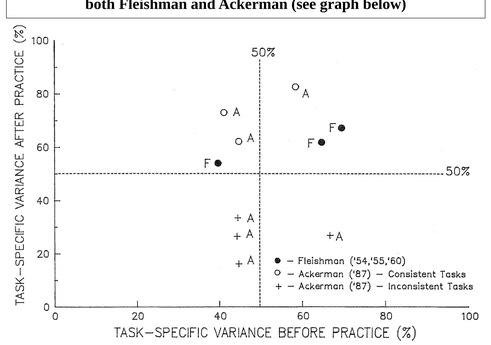
Task-specific variance, before and after practice, compared for consistent (stimulus and response compatible) versus inconsistent (stimulus and response incompatible) tasks

Summary of Findings

- For three consistent tasks, task-specific variance increased from the range 41 to 58 percent before practice, to the range 62 to 83 percent after practice
- For four inconsistent tasks, task-specific variance decreased from the range 44 to 67 percent before practice to the range 18 to 34 percent after practice.

Conclusions

Task-specific variance exceeds 50% after practice (and is between 40 and 50% before practice) for consistent tasks, based on research by both Fleishman and Ackerman (see graph below)



13 Analysis of Individual Differences by Jones (1966) [13]

- This analysis reviewed previous studies of the determinants of sources of variability in cognitive and other learning tasks
- Numerous studies reviewed were found to confirm the work of Woodrow [11], Fleishman [12] and others, showing that the amount of specific variance associated with most cognitive learning tasks is even greater than that found in these original studies

Key Conclusion

"In the case of actual work performance, the variance attributable to the specific design of the task situation is even greater than that found for experimental test situations, amounting usually to at least 75 percent of the total variance, frequently to as much as 90 percent"

14 Usability Paradigm of Gould (1990) [14]

- The conventional assumption in psychology about human-design interaction is reflected in the view of Poulton [15, p. 178)]:

"The aim of engineering psycho1ogy is . . .to specify the capacities and limitations of the human, from which the choice of the better design should be deducible direct1y"

- Gould [14, p. 762] contradicts this view with a context specificity perspective by noting that:

"You Can't Rely Upon Descriptive Data. Even if a description of the intended users were as complete as it possibly could be, it would not be an adequate basis for design. It would not substitute for direct interaction and testing"

Conclusions

Gould's insight is that because variability in human performance is critically referenced to the design of the performance environment (Slides 4-12), the effects of particular designs on user performance cannot necessarily be predicted in advance. Instead, these effects must be established through direct testing.

Because of this insight, Gould now is considered the father of usability testing

Widespread adoption of usability testing (now considered essential for achieving both product and software interface quality) represents a direct manifestation of the replicability problem. That is, designers are a poor judge of how users will actually perform during interaction with their designs

15 Educational Ergonomic Analysis of Smith [16,17]

- Educational ergonomics is defined as that branch of ergonomics/ human factors concerned with the interaction of educational performance and educational design
- The analysis of Smith has focused upon identifying design factors in K-12 learning environments shown to influence academic performance of students enrolled in these classrooms
- A summary of these factors is provided below

Design Factors Whose Possible Effects On Student Performance In K-12 Classrooms Have Been Investigated

Category	Design factor	Impact on student performance
Classroom and school building design Factors	Environmental design of classroom and building facilities Classroom technology Online learning environments	Student academic performance influenced by level of class-room and school building design quality Equivocal effects of computer use No systematic analysis yet available
Educational system Design factors	Smaller class size Longer exposure to learning School choice School funding School size School start times Levels of teacher training and teacher pay	Positive for lower, but inconclusive for higher, grades Strongly positive Varied effects, depending on country No relationship Varied results Non-academic benefits, but academic impact unproved No relationship
Learning strategy Design factors	Cooperative Learning Early Childhood Education Amount of homework Teaching quality	Strongly positive Strongly positive Varied results Necessary, but not sufficient
Student self-confidence	Student beliefs in their learning competence, and in their level of control over the learning context	Varied effects
Design factors influencing student health and wellbeing	Nutritional adequacy Good physical fitness levels and participation in physical activity Emotional well-being	Mostly positive, but not definitive Largely positive Strongly positive
Community system design factors	Informal learning Community socioeconomic status and school-community integration	Promising but limited and not definitive positive results Strongly positive

16 Implications of Educational Ergonomics for Replicability

- Educational ergonomic analysis shows that a variety of ergonomic design factors in learning environments have the potential to influence academic performance (Slide 15)
- Every classroom, school, school system, and school system community has different design features
- Consequently, the specific variance associated with learning performance by students in a given classroom is likely to be considerable
- Conventional theories of learning --- behaviorism, cognitivism, constructivism, emotional, social, take your pick --- each assume a 'one size fits all' understanding of student learnin
- In contrast, evidence for specific variance in behavior and performance (Slides 5-14) suggests that the design of learning environments and strategies should be customized to account for variability in learning behaviors among different students
- With its new Education Initiative, the Bush Foundation (St. Paul, MN) is endorsing this idea [18]

17 Significance of this Analysis

- Evidence reviewed above supports and extends the conclusion of Van Bavel et al. [5] that replicability in psychological research may be compromised by specific variance in behavior and performance.
- Arguably however, the most profound implication of the role of context in replicability is the likelihood that research findings based on laboratory designs may have minimal applicability to real world contexts — such as actual learning environments — with dramatically different designs (Slides 15-156)
- A good example of a research design aimed at avoiding this problem is the study by Johnson et al. [19] investigating the benefits of cooperative learning designs for student attitudes and academic achievement
- The study was carried out in an actual classroom with actual students as subjects, and with the teacher present and assisting with the research
- This approach has yielded good replicability as regards demonstrating the benefits of cooperative learning designs [20]

18 References

1. Diener, E., & Biswas-Diener, R. (2017) . *The Replication Crisis in Psychology* (Universities of Utah and of Virginia, and Portland State University). (http://nobaproject.com/ modules/the-replication-crisis-in-psychology)
2. Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716-1 - aac4716-8.
3. Carpenter, S. (2012). Psychology's bold initiative. *Science, 335,* 1558-1561.
4. Ioannidis, J.P.A. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science, 7*(6), 645-654.
5. Van Bavel, J.J., Mende-Siedleckia, P., Bradya, W.J., & Reineroa, D.A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Science, 113*(23), 6454-6459.
6. Thorndike, E.L. & Woodworth, R.S. (1901a). The influence of improvement in one mental function upon the efficiency of other functions. (I.). *The Psychological Review, 8*(3), 247–261.
7. Thorndike, E.L. & Woodworth, R.S. (1901b). The influence of improvement in one mental function upon the efficiency of other functions. II. The estimation of magnitudes. *The Psychological Review, 8*(4), 383-395.
8. Thorndike, E.L. & Woodworth, R.S. (1901c). The influence of improvement in one mental function upon the efficiency of other functions. III. Functions involving attention, observation and discrimination. *The Psychological Review, 8*(6), 553–564.
9. Poffenburger, A.T. (1915). The influence of improvement in one mental process upon other related processes. *Journal of Educational Psychology, 6,* 459–474.
10. Dewey, J. (1916). *Democracy and education*. New York: Macmillan.
11. Woodrow, H. (1939). Factors in improvement with practice. *Journal of Psychology, 7,* 55-70.
12. Fleishman, E.A. (1966). Human abilities and the acquisition of skill. In E.A. Bilodeau (Ed), *Acquisition of Skill* (147-167). New York: Academic Press.
13. Jones, M.B. (1966). Individual differences. In E.A. Bilodeau (Ed.), *Acquisition of Skill* (pp. 109–146). New York: Academic Press.
14. Poulton, E.C. (1966). Engineering psychology. *Annual Review* of *Psychology, 17,* 177-200.
15. Gould, J.D. (1990). How to design usable systems. In M. Helander (Ed.), *Handbook of Human-Computer Interaction* (pp. 757-789). Amsterdam: Elsevier.
16. Smith, T.J. (2007). The ergonomics of learning: educational design and learning performance. *Ergonomics, 50*(10), 1530–1546.
17. Smith, T.J. (2013). Designing learning environments to promote student learning–ergonomics in all but name. *Work: A Journal of Prevention, Assessment & Rehabilitation, 44,* S39–S60.
18. Reedy, J.F. (2017). Needed: customized schools and curricula. *Star Tribune* (March 3), A9.
19. Johnson, R., Johnson, D.W., & Tauer, M. Effects of cooperative, competitive, and individualistic goal structures on students' achievement and attitudes. *Journal of Psychology,* 1979, *102,* 191-198.
20. Johnson, D.W., Maruyama, G,, Johnson, R,, Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin,* 89, 47-62.