

Evidence-Based Practice and the Reproducibility Crisis in Psychology

Volume 45 Issue 5

By Steven R. Shaw & Joseph D'Intino

Evidence-based practice (EBP) is the norm and expectation for providing interventions in school psychology. Moreover, EBP is considered as best practice by the American Psychological Association, American Academy of Pediatricians, American Speech and Hearing Association, and several other professional associations directly related to care of children, and it has also been codified as a legal requirement in special education under IDEA (2004). However, there are two major hurdles to be addressed before EBP can be a true improvement in providing educational and psychological interventions to children and families. The first challenge is that the standard of clinical research limits the generalizability or application of findings to any given situation as nearly all research, including clinical research, is not replicated or reproduced; which limits the implementation of any intervention that might be based upon that research (Coyne, 2016). A second challenge emerges from the first, as clearer standards need to be established regarding the implementation of EBP to allow for the proper integration of research, clinical judgment, client experience and need, and local or cultural constraints. These challenges can be addressed through a concerted effort and partnership by researchers and clinicians with the goal of improving child and family outcomes. The first steps towards doing so are defining new standards for reproducibility and replication in clinical research, as well as EBP, which can then contribute to a balanced implementation approach.

Defining Evidence-Based Practices

Examining various definitions of EBP provide the potential to develop a clearer standard. EBP in healthcare involves the translation of high-quality scientific research into practice to improve individuals' health and well-being (Steglitz, *et al.*, 2015). In comparison, the Canadian Psychological Association (CPA) defines EBP as a process involving the “conscientious, explicit, and judicious use of the best available research evidence to inform each stage of clinical decision-making and service delivery” (Report of CPA Task Force, 2012, p. 7). This approach to clinical practice was first introduced in a seminal article published in 1992 (Guyatt *et al.*, 1992), which represented the formal introduction of a new decision-making paradigm requiring clinicians to learn new skills based on scientific research in order to provide better care to clients. The shift from authoritative decision making, in which decisions are based on intuition, unsystematic observation, and tradition, to empirically based decision making allows clinicians to avoid biases and poorly supported conclusions based solely on experience (Gambrill, 2006; Guyatt *et al.*, 1992). This change first occurred in the medical world and later spread to other fields such as nursing, occupational therapy, psychology, and education (Gambrill, 2006). In psychology, EBP has the potential to influence and significantly improve the quality of services provided by psychologists.

A similar philosophy has been applied to education as well. The recently enacted Every Student Succeeds Act (ESSA; 2015) defines four levels of EBP that have been codified into law. According to ESSA, “strong” evidence refers to showing a statistically significant effect on student outcomes from at least one experimental study. The terms “moderate” and “promising” require evidence from quasi-experimental study or correlational studies that make statistical corrections for selection bias. Federal funding streams strongly favor educational systems that apply strong, moderate, and promising evidence. In addition, there is a fourth category made up of activities that have a research-based rationale, but lack empirical support. This fourth category is only considered acceptable as long as there are “ongoing efforts to examine the effects” on student outcomes.

Despite the growing expectations to use EBP in psychology and education, a major challenge in the use of EBP is that the construct is interpreted in multiple ways for multiple purposes (Rycroft-Malone, Seers, Titchen, Harvey, Kitspon, & McCormack, 2004). The various definitions, including the ESSA guidelines,

imply the use of research to inform practice, but any single definition remains too vague to establish a standard of EBP. As such, the label of EBP has frequently been used as a talisman that means different things to different people. This is evident when examining two of the three major interpretations of EBP.

Weak EBP. The first approach is the weak form of EBP, which is widely used in marketing and advertising of an intervention. The weak form involves selective use of evaluative data, where ambiguous data is interpreted in a manner that supports the practice at hand. Often, when following this approach, information such as testimonials, program descriptions, or case studies is considered sufficient to label a study as EBP. More commonly, poorly designed studies with large effect sizes are used as proof of EBP. Interventions and practices are thus marketed as receiving with something akin to a seal of approval that meets the legal definition of EBP in education or psychology. This seal of approval is then used to promote books, consulting arrangements, manualized interventions, and other practices that are weak in scientific support. This weak form of EBP is one that is not consistent with the original intent of the concept and dilutes the purpose of EBP. The result of the commonly experienced weak form of EBP is that the term has become meaningless to many clinicians, the culture of implementing new and scientifically supported interventions is more resistant to change, and the purpose of improving clinical practice is harmed.

Severe EBP. The severe form of EBP is equally problematic. In this second approach of implementing EBP, research methods used in the supportive published research studies must be implemented with the exact same methodology in clinical settings. The logic is that deviations from the methodologies used in the supporting research studies will result in different, and probably less positive, outcomes. In this severe form of EBP, concepts of treatment fidelity and treatment integrity are paramount concerns. In education, teachers are frequently monitored to ensure that they implement the intervention exactly as described in the supporting research studies. In medicine, physicians are expected to follow the exact same treatment protocols as described in supporting studies. This interpretation of EBPs leads to a criticism of robotic rather than professional implementation of interventions. There is no opportunity to deviate from the protocol or methodologies of supporting studies. This is obviously a problem when the context, needs of the clientele or clients, resources, clinical expertise, the cultural environment of the institution where interventions are being provided, and a host of other variables are not considered in the severe form of EBP. Furthermore, it is unlikely that the sampling, context, and specifics of the presenting problem are exactly the same for the research group as for the clinical group to which the EBP is being applied. Again, the severe interpretation of EBP is not consistent with the original intent of the concept. The severe form of EBP is frequently rejected, and like the weak form, results in an environment where innovation is unlikely to take place.

Given the challenges associated with EBP, it is not surprising that education has not fully embraced scientific methods and evidence despite laws requiring EBP. Leaders in education have encouraged teachers to “keep the scourge of scientism out of schools” (Furedi, 2013). The idea of using strong research methods such as randomized control trials has been rejected by some as “belonging to discredited view of science as positivism” (Cohen, Manion, & Morrison, 2013, p. 318). The general idea is that the variables required to understand and develop effective interventions in education are too complex and too individualized to children to be effectively characterized by EBPs. However, medical practitioners have likewise rejected aspects of EBPs as that of “cookbook medicine” and “just because something new has been discovered, tested, and found to ‘work well’ does not mean that it is correct.” The entire definition of the concept of EBP focuses on ensuring that the needs of the client be met. In medicine, this requires an assessment of what the client wants or needs (e.g., pain relief, disease prevention, medication, behavior and emotion management techniques, better diagnosis and prognosis). In comparison, psychology requires an assessment of the client's cognitive abilities, emotional state, adaptive and daily living skills, social interactions, and other factors. In education, this requires the determination of academic skills, social and emotional skills, community transition skills, and other factors. EBP is not designed to be robotic or a fig leaf used to promote practices with minimal support, but has its beginning in meeting the needs of clients and students.

A third approach. Both the severe and the weak interpretation of EBP create harm and an environment in which innovation is less likely to be implemented. Consequently, a third, more balanced approach that advocates the integration of research, clinical expertise, client needs, and cultural context is necessary to establish a new and more effective standard of EBP. As outlined in the ESSA, research is the starting point for

such an approach, but a strong research base must be available in order to properly inform practice. Therefore, the first steps towards better EBP and informed practice is to refine the way in which clinical research is conducted and applied.

Problems With Clinical Research

Clinical research plays a critical role in EBP as it is required to inform practice. However, the current standards and culture of clinical research do not lend themselves easily to the application of EBP (Stephens, n.d.). If the evidence from the research is not based upon a robust and valid evaluation of important research questions, then it is challenging to apply the findings into practice and might contribute to the reticence or resistance that some psychologists or educational professionals have towards EBP.

The important challenge for delivery of effective educational, medical, rehabilitation, and psychological services interventions in schools is the implementation of EBPs. Whether information learned from research is true or not is often irrelevant unless it leads to improvement in methods of service delivery (Coyne, Cook, & Therrien, 2016). There is often criticism from clinicians that much of the research in these fields can be called “blue sky research”, which is basic or theoretical research without any obvious or immediate practical implication. Equally, there is criticism from researchers that clinicians are too slow to implement innovative ideas developed in laboratories (Ioannidis, 2016). The concept of “clinical research” has not sufficiently bridged the gap between research and clinical innovation. Clinical research is designed to address questions of treatment, diagnosis and screening, prevention, prognosis, rehabilitation, and other actions with direct benefit to students and clients. Despite large and increasing emphasis placed on clinical research in all fields, there is little evidence that such research has led to any improvement in implementation over blue sky research (Ioannidis, 2016). To a large degree, there is no evidence that implementation of innovation based on clinical or blue sky research is dramatically advancing—especially in areas of education and psychology (Ioannidis, 2016; Klein & Knight, 2005; Lavant & Hasan, 2008).

As explained in the severe approach to EBP, most methodologies or interventions are not generalizable to a wide variety of contexts. Clinical research often provides an initial result or finding that can spark new ideas for innovation. Often, EBP is determined by a single peer-reviewed article that reports that the intervention produced a significant result. As research publication and grants frequently rewards novelty, the pursuit of future directions and questioning suggested by the original study are often abandoned (Shadish, Zelinsky, Vevea, & Kratochwill, 2016).

Several features can be integrated into clinical research to make it useful and easily applicable for clinicians (Ioannidis, 2016). The most important feature is that it should be problem-based and client-centered. This is echoed by the American Psychological Association (2006) as they indicate that treatment efficacy and clinical utility should be the defining elements of clinical research. Other characteristics outlined include that the research be placed within a clinical context, that it contributes to a gain of information, and that it is pragmatic, valuable, feasible, and transparent (Ioannidis, 2016). However, most clinical research fails by design as it rarely features any of these characteristics (p. 2). Studies should therefore be designed with the purpose of improving client experience, and these studies should be evaluated based on their impact and their influence on clinical practice (p.2). In the event that a study should provide a high level of clinical utility, research to further validate those singular findings is necessary.

Perhaps one of the biggest problems associated with current clinical research is the lack of a continuous thread of related research, combined with a lack of replicated findings contributes to a culture of “one and done” research (Shadish *et al.*, 2016). Consequently, practitioners have limited knowledge to inform practice and will most likely rely more heavily on clinical experience and measures without a solid scientific foundation.

The Reproducibility Crisis

Replicated findings contribute to the validity of scientifically supported and EBP. However, a major issue with the concept of EBPs is that basic and applied science often results in extremely small effects that cannot be replicated (Shadish *et al.*, 2016). This issue has been labelled as the reproducibility crisis and has been acknowledged by scientists from a variety of scientific domains (Baker, 2016). Although the problem has been most widely reported and considered to be a problem in the field of psychology (Open Science Collaboration, 2015), there is evidence that problems and reproducibility are common in fields of genetics, medical practices, education, and rehabilitation (Baker, 2016). As such, the terms replication and reproducibility take on different meanings for different contexts. It is important to define reproducibility and replication and properly distinguish them from each other to better identify the problem at hand (Patil, Peng, & Leek, 2016). For example, in psychology, replication involves performing the same experiment twice, matching the participants, data collection, and data analysis with the original study (Irizarry, Peng, & Leek, 2015). In comparison, reproducibility involves the data collection performed by a separate team, on a separate occasion with a different sample, all the while following the same procedure from the original study (Irizarry, Peng, & Leek, 2012). Neither form is common in education or psychology research, but both have the potential to provide further validity and reliability to the research (Biesta, 2007). However, reproduction has the potential to eliminate biases and improve generalizability as the research is carried out with separate samples and different contexts.

The contributing factors to the reproducibility crisis are complex and controversial. Some of the problems revolve around slavish devotion to the concept of statistical significance or $p < .05$. Statistical significance is a function of sample size and the size of the effect, and there is always the possibility that an observed effect would have occurred due to sampling error alone. Moreover, findings can be statistically significant but have no importance for clinical practice. This is the case with large sample studies that find that an innovation results in a statistically significant positive change, but the effect size is small. This means that the innovation has resulted in a just barely noticeable, yet not practically important, change. In addition, many professional scholarly journals have a bias towards including studies with statistically significant results and do not often publish studies that do not show change due to the intervention. As such, only positive results are published, therefore misleading readers into believing that such an innovation has a strong evidence basis. The crisis becomes more complex as the positive results that were published based on small effect sizes are rarely replicated or reproduced. In the event that the effort has been made, the results are often unlikely to produce significant results and, therefore, are not published. In contrast, reproduced findings that do not show significant results do not necessarily imply that the original findings are invalid, but rather that context and design need to be approached more critically (Open Science Collaboration, 2015). Reproduced studies are therefore useful regardless of their statistical significance as they offer insight into the potential influence of different contexts or sample differences. Consequently, they add to the generalizability of the overall research.

Across the various disciplines, scientists recognize the challenge of generalizing findings. In addition, the incentives for reproducing results is low; there is pressure to publish, which requires a focus on novelty and therefore does not motivate researchers or universities to reproduce past experiments (Baker, 2016). Nevertheless, clinical research is designed to address questions of treatment, diagnosis and screening, prevention, prognosis, rehabilitation, and other actions with direct benefit to clients and clients. Reproducing and replicating experiments would further validate research that could be used as a basis for EBP as educational innovations with the strongest level of research support are considered evidence-based (Colquhoun *et al.*, 2014). However, without a strong base of research, it remains difficult to find the proper balance between the weak and severe approaches to EBP. Recognizing that one result is not prescriptive, but rather, that it is a basis for a longer stream of questioning.

Suggestions for School Psychologists

School psychologists can bridge the gap between research and practice in a bilateral manner. Consequently, they have a unique responsibility of influencing the standards for both research and practice. Although the implementation of EBP provides a significant upgrade from interventions based on tradition, mythology, experience, and other unsupported beliefs, EBP remains a difficult concept in practice and a challenge for many professionals, including school psychologists. In order to properly implement an EBP, treatment fidelity

must be considered and followed by school psychologists. Treatment fidelity refers to the implementation of an evidence-based practice as being as close to the methodologies used in the original study providing support for the intervention. Often, treatment fidelity requires direct insertion of an EBP into an existing system with no flexibility in the methods of implementation (Klein & Knight, 2005). As a result, the implementation of EPBs can range from challenging to near impossible.

Theoretically, treatment fidelity could provide excellent guidelines for implementation. However, EBPs need to be frequently adapted to meet the resources (e.g., time, funding, training), cultural issues of the client, systemic needs of the location in which the intervention is implemented, values of the clientele and their families, requirements of sustainability, regulations and laws; yet, also achieve treatment fidelity. A large body of empirical literature suggests that the effectiveness of an innovation is maximized when implemented with fidelity and that deviations from the treatment manual lead to poorer outcomes (Durlak & DuPre, 2008). However, this notion of absolute treatment fidelity is debatable, as programs with treatment adherence rates as low as 40% have been shown to produce positive study outcomes (p.334).

Given the challenges associated with EBP implementation, there are six primary steps required for the effective integration of EBP (CPA Task Force on Evidence-based Practice of Psychological Treatments, 2012; Klein & Knight, 2005; West, n.d.):

- Develop the clinical question using the needs of the clientele.
- Identify all available evidence, including both research and case information, needed to address the clinical question.
- Evaluate the quality and relevance of available scientific evidence for the case at hand.
- Identify contextual and local constraints that might challenge the integrity of the intervention before implementation. Such factors include available resources, cultural barriers, sustainability, cost-efficiency, among others (Ioannidis, 2016).
- Implement the scientifically supported practice.
- Once implementation is complete, evaluate the effectiveness of the intervention for your specific client. Upon evaluating the intervention, specific questions, comments, or observations can be brought forward to the research community to further encourage the inquiry into program effectiveness.

The steps toward effective implementation of EBP can be overwhelming, but they offer a balanced approach to EBP as the intervention is evaluated before, during, and after administering it to the client. For one, identifying effective studies and determining exactly how close those studies match to the problems that have been identified and assessed by clinicians is an extraordinarily taxing and difficult activity. Expecting busy clinicians with large caseloads to conduct literature searches and evaluations of research for each case is not realistic. Cochrane.org and the What Works Clearinghouse (www.w-w-c.org) are examples of easy to access services that evaluate claims of effectiveness in medicine/rehabilitation and education, respectively. These services provide excellent value in determining (a) which interventions have had the best positive outcomes and (b) the quality of the evidence supporting those positive outcomes. However, the onus is still on the clinician to determine whether the EBP best matches the needs of the clientele. Knowing what works is essential, but considering how to apply the EBP is equally, if not more, important for implementation of EBPs. The clinical expertise of the practitioner is critical in the implementation and evaluation process. Clinical expertise refers to the clinician's cumulated experience, education, and clinical skills. The client brings to the encounter his or her own personal preferences and unique concerns, expectations, and values. Similarly, the research brings forward its own strengths and weaknesses and the clinician needs to find the balance between the research and the client's needs.

As a final suggestion, more communication among colleagues regarding recent practices can aid in strengthening reflective practice. Examining, evaluating, and discussing both the research and practice as a team offers an opportunity for clinicians to learn from each other and to discuss the merits and flaws of recent

innovations, client needs, and the balance that should exist therein. Evidence-based practice, when done well, can effectively achieve a balanced approach between research and practice, thereby improving both.

The Future of School Psychology Practice

Evidence-based practice is the current legal requirement of professionals. In education and psychology, better clinical research is necessary for more informed, balanced, and client-focused EBP. Addressing the issues of reproducibility and replication in clinical research can contribute to a stronger base of knowledge that can be helpful and informative for practice. School psychologists play a unique role as bilateral mediators between research and intervention and as such, have a responsibility to apply a balanced approach to EBP and to evaluate effectiveness to inform future research and practice. The days of focusing on the false dichotomy of research versus practice are over. In the age of EBP, generalizable research and implementation to practice as a seamless process is the future of school psychology. In taking advantage of this opportunity to refine the standards of research and practice, better services can be offered to patients, clients, and students.

References

- APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist*, 61(4), 271–285. doi:[10.1037/0003066X.61.4.271](https://doi.org/10.1037/0003066X.61.4.271)
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604), 452–454. doi:[10.1038/533452a](https://doi.org/10.1038/533452a)
- Biesta, G. (2007). Why “what works” won't work: Evidence-based practice and the democratic deficit in educational research. *Educational Theory*, 57(1), 1–22. doi:[10.1111/j.1741-5446.2006.00241.x](https://doi.org/10.1111/j.1741-5446.2006.00241.x)
- Cohen, L., Manion, L., & Morrison, K. (2013). *Research methods in education*. New York, NY: Routledge.
- Colquhoun, H., Leeman, J., Michie, S., Lokker, C., Bragge, P., Hempel, S., ... Grimshaw, J. (2014). Towards a common terminology: A simplified framework of interventions to promote and integrate evidence into health practices, systems, and policies. *Implementation Science*, 9, 781. doi:[10.1186/1748-5908-9-51](https://doi.org/10.1186/1748-5908-9-51)
- Coyne, J. C. (2016, July 6). *A bad response to the crisis of untrustworthiness in clinical psychology science*. Retrieved from <https://jcoynester.wordpress.com/2016/07/05/a-bad-response-to-the-crisis-of-untrustworthiness-in-clinical-psychology-science/>
- Coyne, M. D., Cook, B. G., & Therrien, W. J. (2016). Recommendations for replication research in special education: A framework of systematic conceptual replications. *Remedial and Special Education*, 37, 244–253. doi:[10.1177/074193251932516648463](https://doi.org/10.1177/074193251932516648463)
- CPA Task Force on Evidence-based Practice of Psychological Treatments. (2012). *Evidence-based practice of psychological treatments: A Canadian perspective*. Canadian Psychological Association. Retrieved from http://www.cpa.ca/docs/File/Practice/Report_of_the_EBP_Task_Force_FINAL_Board_Approved_2012.pdf
- Durlak, J. A., & Dupre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and factors affecting implementation. *American Journal of Community Psychology* 41, 327–350.
- Every Student Succeeds Act (ESSA). Retrieved from U.S. Department of Education. (n.d.), <http://www.ed.gov/essa?src=rn>
- Furedi, F. (2013). *Keep the scourge of scientism out of schools: Why evidence-based teaching methods are a bad idea*. Retrieved from http://www.frankfuredi.com/site/article/keep_the_scourge_of_scientism_out_of_schools
- Gambrill, E. (2006). Evidence-based practice and policy: Choices ahead. *Research on Social Work Practice*, 16(3), 338–357. doi:[10.1177/1049731505284205](https://doi.org/10.1177/1049731505284205)
- Guyatt, G., Cairns, J., Churchill, D., Cook, D., Haynes, B., Hirsh, J., ... Tugwell, P. (1992). Evidence-based medicine: A new approach to teaching the practice of medicine. *Journal of the American Medical Association*, 268(17), 2420–2425. doi:[10.1001/jama.1992.03490170092032](https://doi.org/10.1001/jama.1992.03490170092032)

Individuals with Disabilities Education Act, 20 U.S.C. § 1400 (2004).

- Ioannidis, J. P. A. (2016). Why most clinical research is not useful. *PLOS Med*, 13(6), e1002049. doi:[10.1371/journal.pmed.1002049](https://doi.org/10.1371/journal.pmed.1002049)
- Irizarry, R., Peng, R., & Leek, J. (2012, Apr 18). *Replication, psychology, and big science*. Retrieved from <http://simplystatistics.org/2012/04/18/replication-psychology-and-big-science>
- Klein, K. J., & Knight, A. P. (2005). Innovation implementation overcoming the challenge. *Current Directions in Psychological Science*, 14(5), 243–246. doi:[10.1111/j.0963-7214.2005.00373.x](https://doi.org/10.1111/j.0963-7214.2005.00373.x)
- Levant, R. F., & Hasan, N. T. (2008). Evidence-based practice in psychology. *Professional Psychology: Research and Practice*, 39(6), 658–662. doi:[org/10.1037/0735-7028.39.6.658](https://doi.org/10.1037/0735-7028.39.6.658)
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716–aac4716. doi:[10.1126/science.aac4716](https://doi.org/10.1126/science.aac4716)
- Patil, P., Peng, R., Leek, J. T. (2016). A statistical definition for reproducibility and replicability. doi:[org10.1101/066803](https://doi.org/10.1101/066803)
- Rycroft-Malone, J., Seers, K., Titchen, A., Harvey, G., Kitson, A., & McCormack, B. (2004). What counts as evidence in evidence-based practice? *Journal of Advanced Nursing*, 47(1), 81–90. doi:[10.1111/j.1365-2648.2004.03068.x](https://doi.org/10.1111/j.1365-2648.2004.03068.x)
- Shadish, W., Zelinsky, N., Vevea, J., & Kratochwill, T. (2016). A survey of publication practices of single-case design researchers when treatments have small or large effects. *Journal of Applied Behavioral Analysis*. Advance online publication. doi:[10.1002/jaba.308](https://doi.org/10.1002/jaba.308)
- Stephens, J. (n.d.). *Naturalising education: What constitutes as “evidence” in evidence-based educational practice?* Retrieved from <http://naturalisingeducation.blogspot.com/2016/06/what-constitutes-as-evidence-in.html>
- Steglitz, J., Warnick, J. L., Hoffman, S. A., Johnston, W., & Spring, B. (2015). Evidence-based practice. In J. D. Wright (Ed.) *International encyclopedia of the social and behavioral sciences*, (2nd ed.). 332–338. Oxford, UK: Elsevier.
- West, M. R. (n.d.). *From evidence-based programs to an evidence-based system: Opportunities under the Every Student Succeeds Act*. Retrieved from <http://www.brookings.edu/research/papers/2016/02/05-evidence-based-system-opportunities-under-essa-west>
-

STEVEN R. SHAW, PhD, NCSP, is associate professor and graduate program director. JOSEPH D'INTINO is a graduate student in the school and applied child psychology program at McGill University, Montréal, QC

National Association of School Psychologists