OPEN ACCESS Freely available online

ISSN 2737-5315
Volume 3, Issue 3
https://www.iikii.com.sg/journal/EIET
Educational Innovations and Emerging Technologies

Perspective

Toward a Framework for Robust Design-Based Research

M. Shane Tutwiler [1,*], Denise M. Bressler [2], Joseph Reilly [3], Eileen McGivney [4], Tina A. Grotzer [5], Chris Dede [5]

[1] College of Education, University of Rhode Island, Kingston, RI, 02881, USA;
[2] College of Education, East Carolina University, Greenville, NC, 27858, USA; bresslerd19@ecu.edu
[3] College of Professional Studies, Northeastern University, Boston, MA, 02115, USA; j.reilly@northeastern.edu
[4] College of Art, Media & Design, Northeastern University, Boston, MA, 02115, USA; e.mcgivney@northeastern.edu
[5] Graduate School of Education, Harvard University, Cambridge, MA, 02138, USA; tina_grotzer@harvard.edu, chris_dede@harvard.edu
* Correspondence: shane_tutwiler@uri.edu

Received: Apr 24, 2023; **Revised:** Aug 31, 2023; **Accepted:** Sep 10, 2023; **Published:** Sep 30, 2023

Abstract: Design-based research (DBR) is a popular approach for studying and maximizing the effectiveness of learning environments in the Learning Sciences. This approach has historically been approached from a mixed-methodological perspective. In this article, we argue that, with an ever-increasing focus on using the results of DBR to inform policy and practice, the design of DBR studies must be made more robust by addressing issues inherent to the quantitative methodologies employed to track gains in learning. We propose four key design principles (Measurement Matters, Learning is Longitudinal, Use Samples Smartly, and Invest in Fidelity), as well as an analytic framework within which to apply them. A brief case study is used to demonstrate some of these elements in practice.

Keywords: Design-based Research, DBR, Educational technology

1. Introduction

Dede (2004) once asked, “If design-based research is the answer, what is the question?” At the time, design-based research (DBR) was gaining popularity in the Learning Sciences, particularly with individuals focused on researching the design and implementation of computer-supported collaborative learning environments. Dede’s (2004) friendly critique was on the notion that DBR was capable of doing many things, from design and implementation testing to theory formation and testing, almost simultaneously. Characterized as a mixed-methods approach falling somewhere between ethnography and large-scale quantitative experimentation (Collins et al., 2004), the usage of DBR has continued to increase in lock-step with the rise of the Learning Sciences in the past decade (Sommerhof, et al., 2018). That said, the definition of DBR, and its intended purposes, still seems to be somewhat vague. More recently, Puntambekar (2018) argued that DBR is characterized as the iterative development of a learning environment in an authentic context, via a trajectory of purposeful studies that both generate and test key theories and conjectures and culminate in a larger implementation. We agree with Puntambekar (2018) in principle but have found that the advice given therein does not acknowledge the unique design requirements that the quantitative aspects of mixed-methods research necessitate.

In this article, we define a program of robust design-based research (RDBR) that focuses on the quantitative research aspects common to most DBR studies. We refer to this framework as Robust Design Based Research (RDBR). We highlight the fragility of quantitative research, in light of the many assumptions required for inferences to be valid and useful. We then present four principles representing practices in research design that address particular areas of weakness particular to the use of quantitative methods in DBR. We then propose a quantitative methodological approach, Bayesian multilevel modeling, that allows researchers to incorporate our proposed practices into a unified decision-making framework. We conclude this article with a case study of these principles in action, along with final thoughts.

2. Glass Cannon: Quantitative Research

As mentioned, most DBR studies are mixed-methods in nature (Puntambekar, 2018). Methodological guidance for various types of mixed-methods research abounds (e.g., Creswell, 2011). In this article, we seek to reframe this guidance in the service of DBR, specifically to support future research and provide preliminary evidence to policy and practice-based decision-makers. To begin, the quantitative and qualitative components of a mixed-methods study are not equally robust. Internal validity of deep and careful observations, interviews and collected artifacts are a hallmark of qualitative studies conducted with care (Maxwell, 2012).

Quantitative studies, on the other hand, rest heavily on multiple layers of assumptions (Gelman & Hill, 2006). For example, inferences from a common OLS linear regression model (equivalent to a t-test, ANOVA, ANCOVA, etc.) require the following assumptions to be valid and useable: 1) data are randomly sampled from a defined population, 2) the relationship between the mean values of the predictor(s) and the outcome are linearly related, 3) residuals are equally and normally distributed around said linear trend-line, 4) residuals are conditionally independent (i.e., not clustered), 5) there is no measurement error or bias in any of the manifest or latent variables in the model, 6) the sample size is large enough to reliably detect the effect of interest given an assumed minimum effect size (i.e., there was adequate statistical power), and 7) no individual data points are exerting undue leverage on the modeled trend (Gelman & Hill, 2006; Murnane & Willett, 2010, Gelman et al., 2020).

These assumptions need to be checked or verified every time a linear model (e.g., every t-test, ANOVA, ANCOVA, etc.) is used to check for mean differences between variables of interest. Several assumptions, such as linearity of relationships, independence of residuals, and outliers, can be addressed directly in the statistical modeling process. The rest, however, are functions of the design and execution of the research agenda and cannot be addressed post-hoc. Or, as Light, Singer, and Willett (1990) put it, "You can't fix by analysis what you bungle by design." In essence, the quantitative inferential component of any given mixed-methods study is a glass cannon – it is extremely powerful when used as intended, yet equally fragile and prone to breakage if not used perfectly. With that in mind, we developed the following set of recommendations to make the quantitative facets of mixed-methods DBR studies more robust. We distilled these pillars of best practice through dozens of years of combined experience in DBR studies but acknowledge that they are not a perfect solution for every research project. We also note that, being design-focused, they are entirely prescriptive in nature and must not be treated as a lens to judge the quality of previously conducted DBR projects.

3. Pillars of Robust Design-based Research

3.1. Measurement Matters

One of the stickiest, and most ignored, frailties of quantitative research is that inferences are impacted by the quality of measurement of important variables in myriad ways. For example, measurement error may have a higher impact on statistical power than sample size in moderately sized studies, and a small number of samples may systematically increase the likelihood that the magnitude and direction of detected effects might be biased (Gelman & Carlin, 2014). As such, we recommend that researchers endeavoring to engage in DBR spend appropriate effort and resources in identifying, or designing and psychometrically validating, instruments to measure change in cognitive, affective, and behavioral constructs of interest. Systematic validation of meaningful constructs dates back to at least the middle of the 20th century (Cronbach & Meehl, 1955), and must not be overlooked in DBR. On the contrary, given the small sample sizes during the iterative design phase, it needs to take center stage.

3.2. Learning is Longitudinal

It is a given that the social and socio-cultural aspects of learning are accepted and supported throughout the Learning Sciences. That said, learning is always a longitudinal process that occurs, in large part, within each student. That said, quantitative tracking of learning must not be reduced to the use of pre-post (or worse yet, just post) instruments. Instead, researchers must endeavor to measure changes in knowledge, affect, and behavior repeatedly (using the aforementioned robust measures) over the course of the iterative design/implementation phases of the study, as well as during the culminating pilot study. The benefits of this approach are twofold.

Firstly, fitting such data to individual growth models (multilevel regression models with time clustered by student) allows researchers to detect potential non-linear or discontinuous trends in learning (Singer & Willett, 2003). These non-linear trends indicate points during which student learning is suppressed or enhanced by aspects of the design. Researchers can use this information, triangulated with rich qualitative data, to iteratively improve the learning environment and give valuable information to the teacher or facilitator about the need for additional scaffolding. Secondly, such within-student examinations of learning are always inherently more statistically powerful and precise than between-group studies of single mean effects (Gelman, 2018). For example, consider forty students in an implementation study of a DBR project. Randomly assigning twenty students each to a treatment or control condition (ignoring clustering by class and assuming pre-test scores are used as a control) would result in a statistical power of 0.80 to detect effect size differences of about $d = 0.65$ or greater at an alpha of 0.05. In other words, such a design can only reliably detect large effects in learning, and may systematically over or underestimate the nature of smaller effects. On the other hand, a repeated measures design with the same number of students randomly assigned to two groups (or to an A/B comparison) but analyzing four measurements over time would have a power of 0.80 to detect effects of $d = 0.20$ or greater at an alpha of 0.05. In other words, doubling the number of observed measurements nearly tripled the precision of the study.

3.3. Use Samples Smartly

Strategic selection and use of samples in DBR help to bolster the transferability of findings and give decision-makers a better understanding of the potential limitations of inferences they can make based on findings from a DBR study. Sample selection, when possible, must systematically reflect the demographic and cognitive variability of the population of the students for which the learning environment is being designed. This is accomplished, for example, by purposefully choosing classrooms in rural, suburban, and urban schools. The potential heterogeneity in usage and effects allows researchers to better preliminarily model the potential benefits of the learning environment under study, and give decision-makers guidance as to how its adoption may be accepted by their school and students. Furthermore, researchers conducting DBR studies can strategically phase their implementation studies such that different classes from different teachers are cyclically exposed to the different phases of the design. For example, if a research team is working with two teachers, then the classes from Teacher 1 would test the initial design. After iterative updates, the classes from Teacher 2 would test the next design. This process would continue throughout the iterative design process. Doing so prevents the iterative re-designs from over-addressing the needs of an arbitrary group of students and generates evidence about how design choices made in one class may be implemented in a new class. That is to say, transferability.

3.4. Invest in Fidelity

One final necessary investment of resources in the design of a DBR study is the identification of measures of fidelity of implementation. By identifying clear factors of “proper” use of the designed learning environment, and developing scales to track variability in said factors, researchers can identify ways in which variability in the use of learning environments might be related to differentiation in outcomes of interest. One example of this process is found in the fidelity instrument used by the EcoLearn program team at the Harvard Graduate School of Education (McGivney et al., 2019). Evidence of the quality of implementation from this instrument can be used formatively to identify specific aspects of inquiry-based practices, scientific reasoning, general thinking supports, and the degree to which curricular goals were met by teachers in the classroom. This information can be used to qualitatively frame the success of implementation studies, as well as quantitatively predict variation in learning in the implementation and pilot studies, as described below.

4. Unifying Analytics through Multilevel Modeling

As evidenced by the recommendations above, data produced in DBR studies are (or must be) longitudinal, nested by class and research phase, and often sensitive to issues of measurement. Statistical approaches historically employed by learning scientists are not ideally suited to handle multi-dimensional complexity. Advances in computational efficiency and power, as well as applied statistical analytic methods, have opened doors to approaches better suited to answering questions in the Learning Sciences. One particularly powerful approach to analyzing these types of data is the use of Bayesian multilevel models (Gelman & Hill, 2006; Gelman et al, 2013). The use of multilevel modeling in Learning Sciences research has been suggested in the past (Stahl et al., 2008), though it has not been widely adopted. Multilevel regression models allow researchers to predict variation in continuous, nominal, and count-based outcomes while accounting for the clustering of students by pre-existing conditions such as class or school (Gelman and Hill, 2006). In addition, multilevel models form the basis for individual growth models, which allow researchers to model changes in outcomes of interest over time (Singer & Willett, 2003). Measures from validated fidelity instruments further explain differences in learning and help researchers identify criteria for success in transferring these learning environments into new classrooms. Moreover, a Bayesian multilevel modeling approach allows researchers to aggregate data from multiple trials in a DBR cycle of studies (Gelman et al., 2013; Gelman et al., 2020) and allows researchers to account for measurement error and missing data directly (McElreath, 2020).

5. Examples of RDBR

To illustrate our proposed methods, it is necessary to first discuss a simplified case study, and then give further examples of the approach. For pedagogical purposes, as well as for brevity, we abbreviated the theoretical framing that justified this research and focused our conversation on choices of design and analysis that align with the above recommendations. Further details of the studies involved are found in Bressler & Bodzin (2013), Bressler & Bodzin (2016), and Bressler et al. (2019). We explored how the use of an augmented reality-based science curriculum, School Scene Investigators, in a middle-grade science setting might support hypothesized links between engagement and interest (Renninger et al, 2018). Students worked collaboratively to solve two mysteries by scanning QR codes strategically placed in their school and working through a game-based curriculum. While playfully learning, students gathered the required evidence and tried to solve the mystery at hand.

This DBR study highlights four of the pillars of RDBR as follows.

- Measurement Matters
  The flow instrument used in the study was validated and employed in previous research (Jackson, Eklund, & Martin, 2010). In addition, the validity of the instrument in the present data was confirmed via CFA, and instruments were tested for and found to be reliable, with the assessment in each phase demonstrating a Cronbach's alpha score of 0.80 or greater.
- Learning is Longitudinal
  The scientific practice underscores that indexed student learning is developed by collecting and evaluating artifacts of student learning over time, thus capturing the longitudinal nature of student conceptual change. Details of the procedures used to derive the measure are found in Bressler & Bodzin (2016).
- Use Samples Smartly
  For the purpose of this case study, we examined three studies conducted in a broader DBR cycle. The first was a small-scale study (n = 68) testing key design elements in an after-school setting with 6th-8th grade students. The second was a larger scale-up study (n=208) testing the robustness of design changes with 8th-grade students in a different school. The final study (n = 110) was conducted a year later in the same school as the second with a new group of students. By testing across different grade levels and schools, the design changes and detected relationships better captured the heterogeneity common to broaden the implementation of innovations.
- Invest in Fidelity
  Researchers were on-site for each of the implementation studies to ensure that the augmented reality curriculum was used as designed. Researchers tracked any deviations or issues, such as technology failures, which might impact the efficacy of the intervention. As a result, researchers were able to better understand some of the variation in results, as seen below.

In addition to the aforementioned design principles being adhered to, we analyzed the results of each study using Bayesian multilevel logistic regression models (Gelman & Hill, 2006). Based on the results, we modeled the relationship between student engagement/flow and their self-reported gender in a way that maximized out-of-sample predictive power. Figure 1 presents the model estimated relationship between gender and flow across each of the studies, as well as a model that aggregates the findings across all of the studies. The relationship was generally positive, with the studies with higher sample sizes having more precise estimates of the relationship. When looking across all of the studies, on average, girls reported flow scores slightly over 0.2 units of standard deviation (SD) higher than their peer boys.

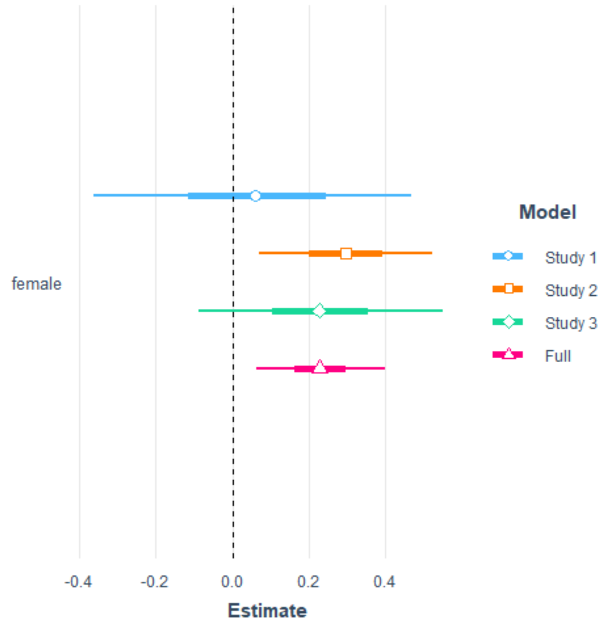

Fig. 1. Model estimated relationship between gender (girl = 1, boy = 0) and flow scores for students engaged in an augmented reality-enhanced science curriculum across three iterative DBR studies. Shapes are median posterior estimates; thin lines are 90% Credible Intervals; thick lines are 50% Credible Intervals. All models were fitted using Bayesian multi-level models.

It is important to note that the pillars of RDBR have been applied more broadly, as well. For example, various RDBR principles can also be found throughout the design, implementation, and analysis of the suite of technology-enhanced science curricula developed by the EcoLearn group at the Harvard Graduate School of Education (Dede et al., 2019). Leveraging various types of immersive virtual environments, the curricula are designed to help teach students about ecosystems and causal patterns in multi-user virtual environments (Grotzer et al., 2013; Metcalf et al., 2018) and via augmented reality (Kamarainen et al, 2018), to include experimentation into their science-based inquiry (Dede et al., 2017), and to help elementary students develop computational thinking skills by studying trends in ecosystems (Metcalf et al., 2021). All of these powerful learning supports were studied iteratively, leveraging the pillars of RDBR in various ways.

- Measurement Matters
  Across the various studies, measures of student learning and affective change were based on previously validated instruments (e.g., Chen at al., 2016) or were rigorously tested for evidence of validity and reliability prior to being used for inference and decision-making (e.g., Tutwiler et al., 2016). This increased the likelihood of detecting important trends in the data and helped to reduce residual variance.
- Learning is Longitudinal
  The design and culminating pilot studies for the various learning scaffolds tracked changes to student learning and attitude over time (e.g., Cuzzolino et al., 2019; Metcalf et al., 2020), and longitudinally modeled changes in student behavior within the immersive learning environments (e.g., Tutwiler, 2019). Mappings of student movement through the virtual environments were also tracked over time (e.g., Courter et al., 2014; Grotzer et al., 2015).
- Use Samples Smartly
  Classes of students were selected from schools that varied on demographic and socio-economic dimensions, allowing the findings of the various final pilot studies to be more broadly generalizable and better guide decision making (e.g., Metcalf et al., 2013; Cuzzolino et al., 2019).
- Invest in Fidelity
  During the design and quasi-experimental pilot studies, researchers were on site to help ensure fidelity of implementation (e.g., Metcalf et al., 2013), and specific rubrics were developed to track fidelity of implementation over time (e.g., McGivney et al., 2019). This allowed for a more meaningful interpretation of the observed effects of implementation over time.

The data from these various studies were analyzed in ways that accounted for their multilevel structure, with students often being clustered by class, teacher, and school (e.g., Kamarainen et al., 2013; Cuzzolino et al., 2019; Tutwiler, 2019).

6. Conclusion

For the past three decades, Learning Scientists have leveraged DBR to explore the design and implementation of learning environments (Brown, 1992). Changes in evidentiary standards in the broader social science research community, largely in light of the replication crisis, have resulted in evolving best practices in quantitative research (Frias-Navarro et al., 2020). In this paper we align DBR with these changes (Fig. 2), resulting in more robust, transferable findings that can help to guide future research and practice. Increasing the focus on measurement error and focusing on longitudinal modeling will increase the power and precision of DBR studies. We refer to this framework as Robust Design Based Research (RDBR).

Choosing diverse samples of students, tracking factors that indicate the fidelity of implementation, and integrating data from across multiple implementation and pilot studies will help researchers to predict what facets of their designs will increase their transfer to other classrooms. In addition, using RDBR design principles will also help researchers to move the findings of their implementation and culminating pilot studies forward in robust, scaled-up studies of efficacy and impact using traditional cluster randomized control-based methods (Murnane & Willett, 2010). Facilitating the scale-up of the study of learning environments will, in the long run, allow stakeholders to make better informed decisions about what might work for their students.

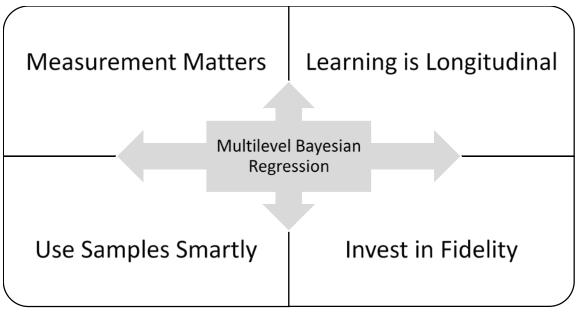

Fig. 2. Robust Design-based Research (RDBR): The four pillars enacted in a multilevel Bayesian framework.

Author Contributions: All authors contributed equally to the conceptualization and writing of this perspective article.

Funding: This research did not receive any external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bressler, D.M., & Bodzin, A.M. (2013). A mixed methods assessment of students' flow experiences during a mobile augmented reality science game. *Journal of Computer Assisted Learning*, *29*(6), 505–517.
2. Bressler, D.M., & Bodzin, A.M. (2016). Investigating flow experience and scientific practices during a mobile serious educational game. *Journal of Science Education and Technology*, *25*, 795–805.
3. Bressler, D.M., Bodzin, A.M., & Tutwiler, M.S. (2019). Engaging middle school students in scientific practice with a collaborative mobile game. *Journal of Computer Assisted Learning*, *35*(2), 197–207.
4. Brown, A.L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, *2*(2), 141–178.
5. Chen, J.A., Tutwiler, M.S., Metcalf, S.J., Kamarainen, A., Grotzer, T., & Dede, C. (2016). A multi-user virtual environment to support students' self-efficacy and interest in science: A latent growth model analysis. *Learning and Instruction*, *41*, 11–22.
6. Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *The Journal of the Learning Sciences*, *13*(1), 15–42.
7. Courter, C.J., Grotzer, T.A., Derbiszewska, K.M., Powell, M., Tutwiler, M.S., Kamarainen, A., Metcalf, S.J., & Dede, C. (2014). A case study contrasting students' exploration behaviors of a complex causal scenario in a virtual world: Mapping proximity and goal directed movement in EcoMUVE. In Proceedings of National Association of Research in Science Teaching (NARST) Annual International Conference. Pittsburgh, PA, USA, 30 March–3 April, 2014.
8. Creswell, J.W. (2011). Controversies in mixed methods research. *The Sage Handbook of Qualitative Research*, *4*, 269–284.
9. Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281.
10. Cuzzolino, M.P., Grotzer, T.A., Tutwiler, M.S., & Torres, E. (2019). An agentive focus may limit learning about complex causality and systems dynamics: A study of seventh graders' explanations of ecosystems. *Journal of Research in Science Teaching*, *56* (8), 1083–1105.
11. Dede, C. (2004). If design-based research is the answer, what is the question? A commentary on Collins, Joseph, and Bielaczyc; diSessa and Cobb; and Fishman, Marx, Blumenthal, Krajcik, and Soloway in the JLS special issue on design-based research. *The Journal of the Learning Sciences*, *13*(1), 105–114.
12. Dede, C., Grotzer, T. A., Kamarainen, A., & Metcalf, S. (2017). EcoXPT: Designing for deeper learning through experimentation in an immersive virtual ecosystem. *Journal of Educational Technology & Society*, *20*(4), 166–178.
13. Dede, C., Grotzer, T., Kamarainen, A., Metcalf, S., & Feldman, R. (2019). Designing immersive authentic simulations that enhance motivation and learning: EcoLearn. *Learning science: Theory, research, practice*, 229–259.
14. Frias-Navarro, D., Pascual-Llobell, J., Pascual-Soler, M., Perezgonzalez, J., & Berrios-Riquelme, J. (2020). Replication crisis or an opportunity to improve scientific production?. *European Journal of Education*, *55*(4), 618–631.
15. Gelman, A. (2018). The failure of null hypothesis significance testing when studying incremental changes, and what to do about it. *Personality and Social Psychology Bulletin*, *44*(1), 16–23.
16. Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science*, *9*(6), 641–651.
17. Gelman, A., Carlin, J.B., Rubin, D.B., Vehtari, A., Dunson, D.B., & Stern, H.S. (2013). Chapman and Hall/CRC. In *Bayesian data analysis*, New York, USA: Taylor & Francis.
18. Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models.* Cambridge, UK: Cambridge University Press.
19. Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and other stories*. Cambridge, UK: Cambridge University Press.
20. Grotzer, T.A., Kamarainen, A.M., Tutwiler, M.S., Metcalf, S., & Dede, C. (2013). Learning to reason about ecosystems dynamics over time: The challenges of an event-based causal focus. *BioScience*, *63*(4), 288–296.
21. Grotzer, T.A., Powell, M., Derbiszewska, K.M., Courter, C.J., Kamarainen, A.M., Metcalf, S.J., & Dede, C.J. (2015). Turning Transfer Inside Out: The Affordances of Virtual Worlds and Mobile Devices in Real World Contexts for Teaching About Causality Across Time and Distance in Ecosystems. *Technology, Knowledge and Learning*, *20*(1), 43–69.
22. Jackson, S., Eklund, R., & Martin, A. (2010). *The FLOW manual*. Queensland, Australia: Mind Garden, Inc.
23. Kamarainen, A.M., Reilly, J.M., Metcalf, S.J., Grotzer, T.A., & Dede, C.J. (2018). Using Mobile Location-Based Augmented Reality to Support Outdoor Learning in Undergraduate Ecology and Environmental Science Courses. *The Bulletin of the Ecological Society of America*, *99* (2), 259–276.

24. Light, R.J., Singer, J.D., & Willett, J.B. (1990). *By design*. Cambridge, MA: Harvard University.
25. Maxwell, J.A. (2012). *Qualitative research design: An interactive approach (Vol. 41)*. Los Angeles, CA, USA: Sage publications.
26. McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan.* New York, NY, USA: Chapman and Hall/CRC Press.
27. McGivney, E., Gonzalez, E., De Los Santos, S., Kamarainen, A. & Grotzer, T.A. (2019). Improving understanding of teaching practice for student learning: A holistic measure of fidelity of implementation. Presented at the National Association for Research in Science Teaching (NARST), Baltimore, MD, USA, 2 April 2019.
28. Metcalf, S.J., Dickes, A.C., Reilly, J.M., Kamarainen, A.M., Brennan, K.A., Grotzer, T.A., & Dede, C.J. (2020). Impact of a Blended Immersive and Computational Modeling Tool on Elementary Ecosystems Science Learning. Available online: https://ecolearn.gse.harvard.edu/node/1477274 (accessed on 15 July 2023)
29. Metcalf, S.J., Kamarainen, A.M., Torres, E., Grotzer, T.A., & Dede, C.J. (2018). Integrating Multi-User Virtual Environments in Modern Classrooms, In *EcoMUVE: A Case Study on the Affordances of MUVEs in Ecosystem Science Education*. Hershey,PA, USA: IGI Global.
30. Metcalf, S.J., Reilly, J M., Jeon, S., Wang, A., Pyers, A., Brennan, K., & Dede, C. (2021). Assessing computational thinking through the lenses of functionality and computational fluency. *Computer Science Education*, *31*(2), 199–223.
31. Murnane, R.J., & Willett, J.B. (2010). *Methods matter: Improving causal inference in educational and social science research.* New York, NY: Oxford University Press.
32. Puntambekar, S. (2018). Design-Based Research (DBR). In *International Handbook of the Learning Sciences*. New York, NY: Routledge Press; pp. 383–392.
33. Renninger, K.A., Ren, Y., & Kern, H.M. (2018). Motivation, engagement, and interest. In *The International Handbook of the Learning Sciences*. New York, NY: Routledge Press; pp. 116–126.
34. Singer, J.D., Willett, J.B., & Willett, J.B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York, NY: Oxford University press.
35. Sommerhoff, D., Szameitat, A., Vogel, F., Chernikova, O., Loderer, K., & Fischer, F. (2018). What do we teach when we teach the learning sciences? A document analysis of 75 graduate programs. *Journal of the Learning Sciences*, *27*(2), 319–351.
36. Stahl, G., & Hesse, F. (2008). The many levels of CSCL. *International Journal of Computer-Supported Collaborative Learning*, *3*(1), 1–4.
37. Tutwiler, M.S. (2019). Exploring The relationship between attentional capture and prior knowledge in a science-based multi-user virtual environment: an individual growth model analysis. *Journal of Science Education and Technology*, *28*, 299–309.
38. Tutwiler, M.S., Grotzer, T.A., Thompson, M.M., Kamarainen, A.M., Metcalf, S.J., & Dede, C.J. (2016). Validation of an instrument measuring student complex causal assumptions. In Proceedings of National Association for Research in Science Teaching (NARST) Conference, Baltimore, MD, USA, 14–17 April, 2016.

Publisher's Note: IIKII stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.