International Journal of Methodology

Volume 1, Issue 1, pp. 2-10, 2022
DOI: https://doi.org/10.21467/ijm.1.1.4668

From Clinical Trial to Education: Methodologies, Assumptions, and Directions

Anna Yinqi Zhang
Department of Educational Psychology, Counseling, and Special Education,
Pennsylvania State University

*Corresponding author email: ybz5148@psu.edu
Received: 11 December 2021 / Revised: 28 January 2022 / Accepted: 18 February 2022 / Published: 08 March 2022

ABSTRACT

In the field of education, emphasis on evidence-based practice, randomized controlled trials (RCTs), causal inference, and process evaluation can all find their roots back in clinical trials and medical research. This response paper surveys contemporary literature in psychometrics, process evaluation, and RCTs aiming to evaluate the feasibility and limitations of RCTs as a methodology in education and provide future directions. Based on the systematic literature review, the author argues: (1) A lack of significant positive treatment effect does not indicate that RCTs are not worth the investment. (2) A careful evaluation of the intervention itself, implementation process, and measurement instrument is recommended for RCTs. (3) There is the need to reframe some causal inference assumptions in an education setting. The paper also provides several examples of reframing assumptions and comments on the caveats. In conclusion, the author foresees a promising future for RCTs in education with the appropriate reframing of assumptions, process evaluation and replication, and recognition of the validity of parallel methodologies.

Keywords: Randomized Control Trials, Education, Causal Inference

1 Introduction

1.1 RCTs in Education and Medicine

In the field of medicine and health care, RCTs are deemed as the most valid means to provide evidence for the effectiveness of a treatment (Cochrane, 1972). The underlying logic of RCTs is that once ensuring initial group equivalence, the randomization process allows us to make causal conclusions of the effect of treatment compared to control/business-as-usual groups. RCT is a valuable experimental method to verify the effectiveness of treatments and compare the effectiveness of two or more treatments. The most recent developments of RCTs are mostly in their applications, as RCTs are one of the well-developed research toolkits for research and development. In recent years, RCTs are widely adopted to test the effectiveness of diseases treatment such as esophageal cancer (Watanabe et al., 2020), osteoarthritis (Grässel & Muschter, 2020), sickle cell diseases (Salinas Cisneros & Thein, 2020), etc.

There has been a push towards evidence-based practice in education since the late 1990s (Pring & Thomas, 2004), primarily to gauge the effectiveness of intervention and programs (Connolly et al., 2018). In education, RCTs are the most widely adopted clinical methodology to support evidence-based programs and examine the effect of interventions on students' developmental and learning outcomes (Connolly et al., 2017; Connolly et al., 2018; Torgerson & Torgerson, 2008). In Special Education, at least two RCTs, four quasi-experimental group comparison studies, or five SCDs are required to show a practice is evidence-based (Council for Exceptional Children [CEC], 2014).

Since its inception, CRT has become widely accepted with institutional support and funding in education worldwide (e.g., the formation of the Institute of Education Sciences (IES) and the passing of the Education Science Reform Act in the United States and the Education Endowment Foundation (EEF) in the United Kingdom). In addition, RCT is viewed as the golden standard for field experimentation and

Copyright © 2022. The Author(s). Published by AIJR Publisher.
This is an open access article under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license, which permits any non-commercial use, distribution, adaptation, and reproduction in any medium, as long as the original work is properly cited.

to produce evidence for empirical research in education (Odom, 2021). McKnight & Morgan (2020) refers to the phenomenon as the 'bullying effect' of RCT, as other non-RCT methods such as quasi-experiments, single-case designs, and qualitative studies have been relegated to the second place. Connolly et al. (2018) reported that 1,017 RCT studies were published between 1980 and 2016 and that 75% of them were published within the last ten years of the aforementioned period. Yet, it is not the case that RCTs have never encountered any methodological concerns or criticisms. Hedges & Schauer (2018) commented on the almost abandonment of the education experiments between the 1980s and early 2000s, with the debates on the relevance of RCTs to the field of education, largely due to a lack of positive effects of the interventions evaluated.

RCT, as a methodology, has been widely adopted in general education and special education to address a diverse body of theoretical questions. RCTs yield state-of-art treatment and objective data on student performance, which many parents appreciate (O'Toole et al., 1998). Education researchers conduct RCTs to study the effectiveness of intervention programs which claimed to reduce the frequency of problematic behaviors in the classroom and increase student-teacher interactions (Sutherland et al., 2018), facilitate job skill acquisition and increase the employment rate for youth with autism (Wehman et al., 2017), improve students' academic achievement and willingness to intervene in bullying situations (Espelage et al., 2016), ease the transition from primary school to secondary schools for students with autism (Mandy et al., 2016), etc. From 2020 to 2022, RCTs have been carried out to investigate the feasibility of a web-based program to promote a healthy diet in early childhood (Barnes et al., 2021), a health literacy program for elderly people (Delavara, Pashaeyp, & Negarandeha, 2020), and many more.

1.2 Comparing Education and Medicine Research

There are many similarities in terms of semantics (e.g., the procedure of posing research questions and implementing intervention programs) in both medical and education research. The latter endeavor, however, tends to have more noise (Blanchard et al., 2014). Blanchard et al. (2014) outlined several differences between clinical and educational research. First of all, clinical researchers are only responsible for measuring the effectiveness of treatments but not for the development of the treatments, but education research usually involves both the innovation and field-testing of the effectiveness of intervention (treatment) (Blanchard et al., 2014). When it comes to measurement and assessment, in clinical settings, researchers measure observable biological and physiological traits, whereas education researchers are often responsible for both creating and using measurement instruments of latent constructs (Blanchard et al., 2014). A latent construct, though not directly observable, is believed to influence how individuals will behave and needs to be operationally defined (Byrne, 2013). Identifying the relevant latent constructs poses a challenge, too. Clinical researchers rely on knowledge of the diseases and human physiological states to identify and measure relevant traits. Education research is much more theory-driven (Blanchard et al., 2014). Theories specify the relationship between constructs and help researchers define the scope and context of their investigation. The acquisition of validity and reliability evidence for an assessment tool must be for the purported use with a particular sample and in a specific context (Rickards et al., 2012). Reliability is a measure of consistency on scores but not a characteristic of the instrument. It is thus incorrect to say that a particular measurement instrument is valid or reliable. Lastly, clinical research aims to isolate the treatment effect, in which a confounding variable is often viewed as 'noise' (a nuisance variable) (Blanchard et al., 2014). In education research, viewing the confounding variable as a nuisance variable dismisses the complexity of dynamic local context investigated, as the confounding variable could be the reasons why an intervention may or may not work (e.g., the effect of the lecture material, characteristics of the teacher & of the students). Yet despite the differences between the two fields, emphasis on evidence-based practice and methodologies such as randomized controlled trials (RCTs), causal inference, and process evaluation (Linnan & Steckler, 2002) can all be traced back to the medical field.

1.3 Research Questions

The paper surveys contemporary literature in psychometrics and education in general relevant RCTs aiming to evaluate the feasibility and limitations of RCTs as a methodology. It aims to offer a comprehensive response to the research question proposed by integrating different perspectives on RCTs, causal inference, and process evaluation. The latter two topics are naturally entailed in the paper, as there is the need for a careful evaluation and reframing of causal assumptions under RCTs and process evaluation with increasingly complex service interventions. The paper responded to several arguments in the Odom (2021) article *Education of Students with Disabilities, Science, and Randomized Controlled Trials* and reported on Hong & Raudenbush (2006) and Raudenbush (2008) causal inference models in detail as a genuine response to the call for a careful examination of the premises of RCTs. The paper aims to address the following research question: How can we best use RCTs, a methodology originating in clinical research, in education? What are the assumptions and caveats?

2 RCTs and Causal Inference: Concerns and Responses

2.1 Main Argument #1. A lack of significant positive treatment effect does not indicate that RCTs are not worth the investment.

Odom (2021) questioned the use of RCTs as the primary methodology for intervention in authentic school settings in both general and special education. The paper made a sweeping argument that RCTs in education are not necessarily worth the investment, citing Lortie-Forgues and Inglis (2009) and Schneider (2019)'s keynote speech at the 2019 IES Principal Investigator Meeting. The former reported that among the RCT research funded by EEF and National Center for Education and Evaluation and Regional Assistance (NCEE), 23% reported a positive effect associated with the intervention, 38% reported a positive effect associated with control/contrast, and 40% results were considered non-informative or lacking the power to make a definite conclusion. The medium effect size reported was 0.03 and the unweighted mean was 0.06, which was fairly small given the standard in Lipsey et al. (2012). This meta-analysis result aligns well with the latter keynote speech that 75% of funded RCT research by IES did not find any significant positive effect for the investigated interventions (Schneider, 2019).

Several published articles address the aforementioned concerns of effect size and the absence of positive effects associated with the intervention. An absence of positive effects associated with the intervention does not mean that a rigorously designed RCT is a failure. Indeed, the absence of a positive intervention effect can be attributed to either the ineffective intervention itself or the implementation failure (Boylan and Demack, 2018). For example, the teachers may not faithfully carry out the intended instructional regimes. RCT as a methodology should not bear the criticism of such implementation failure. It is also possible that the instructional regimes in the control/business-as-usual group substantially overlap with those in the intervention group.

Moreover, rigorous evidence of no or harmful effects should be viewed as valuable scientific knowledge and communicated to practitioners, researchers, funders, and policy-makers so that they can resort to or move on to testing alternative interventions (Styles & Torgerson, 2018). Ironically, poorly designed trials are more likely to find positive results (Styles & Torgerson, 2018). Therefore, education researchers now and then should still embrace rigorous experimental research designs to study the effect of fully developed and replicable education intervention whenever necessary.

2.2 Main Argument #2. A careful evaluation of the intervention itself, implementation process, and measurement instrument is recommended for RCTs.

Odom (2021) commented that RCTs may produce beneficial effects on proximal measures but fail to create long-term beneficial outcomes on the norm-referenced measures. However, I would argue that the absence of effect on distal measures may not be solely attributable to the use of RCTs. We need to carefully examine, along with the rigorously designed RCT procedure, (1) how well the proximal and distal

measures are correlated, (2) how 'distal' is the standard norm-referenced measure, and (3) how well the distal measurement instruments reflect the outcomes of interest of the intervention. As commented in Styles & Torgerson (2018) - 'it is often forgotten that an ability to measure relevant outcomes well is a prerequisite for embarking on an RCT.' The fidelity of implementation and the length of the intervention should be under scrutiny as well, which a deliberate process evaluation can achieve.

Acknowledgedly, there exist types of interventions not agreeable to be tested by randomized controlled trials. Odom (2021) argued that RCTs may not be appropriate for complex service interventions, especially in the cases where intervention needs to be adjusted at a local level (e.g., individualized approach for students with special needs) or where the intervention is received with resistance from school administrations and teachers. Nevertheless, RCTs question the nature of readiness for the evaluation of an intervention. If an intervention is not suitable for evaluation using RCT, it is likely that the interventions to be evaluated may not be fully developed and replicable yet (Styles & Torgerson, 2018). Dawson et al. (2018) made a similar argument by pointing out that some of the EFF-funded interventions were indeed not ready for RCTs, a responsibility shared by the developer, the funder, and the individual evaluator, which can be addressed by more systematic planning and review before the trial.

Another perspective of reframing assumptions is to relax some RCTs' assumptions when supporting evidence is provided. One such supporting evidence is process evaluation on implementation fidelity, participants and social context, participants' appraisal, etc. (Odom, 2021). Similar arguments have been made in Connolly et al (2018), Dawson et al. (2018), Siddiqui et al. (2018), and Styles and Torgerson (2018). As RCTs give us confidence in the isolated effect of the intervention, process evaluation devolves into the underlying reasons of why such an effect occurs. Process evaluation examines which component of the intervention leads to the desirable outcome (Linnan & Steckler, 2002) and derives lessons for future research by helping us understand which RCT results could contradict one another (Siddiqui et al., 2018). Siddiqui et al. (2018) differentiated between fidelity of design and fidelity of implementation. For example, schools overrode initial randomization by switching participants between intervention and control groups out of the belief that disadvantaged participants are more likely to benefit from the intervention (Siddiqui et al., 2018). The phenomenon was recovered during process evaluation, without which it is likely that we would never be made aware of the switching and intent-to-treat analysis would never be done. Sadly, process evaluation is only in its emergent stage of being accepted and used in education: the component of process evaluation is missing in two-thirds of the RCTs reviewed by Connolly et al. (2018).

2.3 Main Argument #3. There is the need to reframe some causal inference assumptions in an education setting.

There are numerous reasons for reframing some of the causal inference assumptions of RCTs. Clinical trials in medicine and health policy have parallels in curriculum and instruction research, but the latter pose their unique challenges given the social nature of the inquiry (Raudenbush, 2008). The social structures of students nested in classrooms and classrooms nested in schools violate the fundamental assumptions of causal inference in a clinical setting (Raudenbush, 2008). Thus, when conducting causal inference of a student's potential academic outcome, we should account for those multiple layers of influences: peers, teachers, classrooms, school, and school districts.

There is also the logistic issue. In the USA, UK, and Scandinavia, teachers are not the common initiators of RCTs (Styles & Torgerson, 2018). Teachers could be informed that they are in the business-as-usual groups and may provide more intensive instruction to compensate for the hypothetical beneficial effect of the intervention assigned. Researchers should take into consideration teachers' attitudes in the classroom when conducting RCTs. Professional development might help to decrease teachers' resistance to participating in RCTs. It is also likely that students could be informed of their experimental group assigned and act accordingly to engage or disengage with the intervention.

2.4 Causal Inference: Uncritical Application Without Reframing Assumptions

As stated before, many concepts in the clinical setting will require a reframing of assumptions when transferring to an education context (Blanchard et al., 2014). Still, let us first examine a case of uncritical application of the causal inference model (without any reframing of assumptions) in education.

First of all, the assumption of initial group equivalence needs to be satisfied in RCTs. Odom (2021) recommended two strategies to achieve group equivalence in small samples prior to RCTs. The first strategy is to match and strategy samples based on factors such as demographic information, socioeconomic status, and cognitive abilities. The goal is to minimize the influence of some unmeasured variables, associated with the matching factors, on the intervention outcomes. The second strategy is to repeat the random assignment process with non-equivalent groups from the initial assignment until between-group equivalence is achieved. The repeating random assignment process to ensure group equivalence addresses the concern in Styles and Torgerson (2018) that randomization tends not to occur within the subgroup and that the subgroup analysis may be prone to 'familywise error rate.'

To write the causal inference model (Holland, 1986) in an educational context: $Z = 1$ indicates that a student receives the instruction intervention and $Z = 0$ indicates that a student receives the business-as-usual instruction. The goal of conducting a RCT is to compare the instructional intervention and standard/business-as-usual instruction on students learning outcome Y. The quantity of interest is expressed as $\Delta = Y(1) - Y(0)$, a student-level causal effect. Holland (1986) termed the phenomenon as 'the fundamental problem of causal inference,' which Δ can never be directly observed. If the student is assigned to the instructional intervention, we can observe $Y(1)$ but not $Y(0)$, the counterfactual outcome. If the student is assigned to the business-as-usual instruction, then we can only observe $Y(0)$ but not $Y(1)$.

Under the two following assumptions, we can estimate the population-average causal effect δ using the equation:

$$\delta = E(\Delta) = E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)] = Y_{1_hat} - Y_{0_hat}$$

Assumption One. Y_{1_hat} is an unbiased estimator of $E[Y(1)]$ and Y_{0_hat} is an unbiased estimator of $E[Y(0)]$, which motivates the need for randomized controlled experiments.

Assumption Two. The Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1986). In a clinical trial setting, SUTVA implies that there will be 'no interference between units' and a patient's outcome will not depend on the physicians who prescribe the drug. Thus, we can rewrite $Y_1(Z_1, Z_2, Z_3, \ldots, Z_n; d) = Y_1(Z_1)$, where $d = (1, 2, 3, \ldots, D)$ is the total number of physicians available. In an educational setting, SUTVA is framed as a student's potential educational outcome is stable regardless of either the instructions or other types of treatments assigned to other students or the instructors/assignment mechanisms. Equivalently, we can rewrite the above equation as $Y_1(Z_1, Z_2, Z_3, \ldots, Z_n; t) = Y_1(Z_1)$, where t represents the instructor available.

Without modification to the assumptions, Raudenbush (2008) pointed out that the SUTVA may not hold for instructions in education. It is very unlikely that a student's potential outcome is only dependent on the treatment assignment itself. One's learning outcome can certainly be influenced by the instructional intervention assigned to other students, educational background and motivation status of other students in the same classroom, and teacher characteristics.

2.5 Causal Inference: Reframing of Assumption for Whole Classroom/School Assignment

To make the SUTVA applicable in an educational setting, Raudenbush (2008) chose to randomly assign the entire classroom or school to treatment intervention. The whole classroom/school assignment approach addresses the concern of Odom (2021): some students are only enrolled in one class. Randomly assigning students in the same classroom to treatment and control conditions could be problematic because teachers may encounter difficulty using experimental intervention with some of the students but not with

the other. To write it in a causal inference language, there are n_j students in each classroom j. The potential outcome of a student particular student $i \in (1, 2, 3, \ldots n_j)$ is expressed as $Y_{ij}(Z_{1j}, Z_{2j}, Z_{3j}, \ldots, Z_{n_j j}; t_j)$. Two more assumptions need to be satisfied for the SUTVA to hold.

Assumption One. There should be no interference between classrooms. For example, teachers assigned to intervention and business-as-usual instruction classrooms should not share information in a way that will influence the students' potential learning outcomes.

Assumption Two. The intact classroom assumption states that students and teachers will not move from one classroom to another. Note that this assumption is likely to be violated if the intervention lasts more than one year, when students could migrate to new classrooms and even new schools.

With the 'no interference between classroom' and 'intact classroom' assumptions satisfied, we are assigning the classrooms as a whole to intervention or business-as-usual groups. We know $Z_{1j}, Z_{2j}, Z_{3j}, \ldots, Z_{n_j j} = 1$ or $Z_{1j}, Z_{2j}, Z_{3j}, \ldots, Z_{n_j j} = 0$ and the potential outcomes becomes $Y_{ij}(1,1,1,\ldots,1; t_j)$ or $Y_{ij}(0,0,0,\ldots,0; t_j)$ because a classroom only receives either treatment or control. The student-level outcome can be expressed as $\Delta_{ij} = Y_{ij}(1,1,1,\ldots,1; t_j)$ - $Y_{ij}(0,0,0,\ldots,0; t_j)$. Raudenbush (2008) argued that the random assignment of classrooms allows for unbiased estimation of the treatment effect and the varying causal effect of Δ_{ij} from school to school and even district to district. Furthermore, Hong and Raudenbush (2006) proposed the estimation of causal effect $Y_{ijk}(1) - Y_{ijk}(0)$, where Y refers to the outcome of student i in classroom j school k with the 'intact school' assumption.

2.6 Causal Inference: Reframing of Assumption: Four-Way Hierarchical Model with Inverse Probability of Treatment Weighting (IPTW)

A student's learning can occur in prolonged periods and multiple social contexts when given a sequence of instructions. The cumulative effect of an instructional regime over many years cannot simply be understood as the sum of instructional effects of each year (Hong & Raudenbush, 2006; Raudenbush, 2008). Beneficial or harmful instructional effects in the later years may largely depend on the instructional effects in the earlier years. In the case of high mobility of classrooms and schools & instructional regimes differing within schools, a phenomenon called 'time-varying confounding' could occur (Raudenbush, 2008).

Let us see how Raudenbush (2008) numerically formulate the problem. We denote Y_0 as the educational outcome at the baseline. $Z_1 \in 0,1$ represents the treatment received by a student in year 1 and the year-end outcome is denoted as Y_1. Similarly, $Z_2 \in 0,1$ represents the treatment received by a student in year 2 and the year-end outcome is denoted as Y_2. While we usually aim to study the effect of Z_1 on Y_1 and the joint effect of Z_1 and Z_2 on Y_2, Y_1 is considered as a 'time-varying confounder' and a causal pathway between Z_1 and Y_2 (Raudenbush, 2008). To understand the joint effect of $Z_1$and Z_2 on Y_2 (e.g., the effect of the sequence of instruction at the end of Year 2), we cannot simply control Y_1. Raudenbush (2008) argued that if we control Y_1, we will bias our estimation of Z_1 on Y_2; if we fail to control Y_1, we will bias our estimation of Z_2 on Y_2.

Robins (2000) proposed the method of inverse probability of treatment weighting (IPTW) to resolve the problem, where Y_1 is used as a weight instead of a covariate. Participants who are less likely to be selected into the treatment will be weighed up and those who are more likely to be selected into the treatment will be weighed down. In scenarios with observed covariates X and the unobservable covariate U, Raudenbush (2008) specified that U_1 should be independent of $Z_1|Y_0, X_1$ and U_1, U_2 should be independent of $Z_2|Y_0, Y_1, X_1, X_2, Z_1$ The two assumptions need to be satisfied before researchers weigh up or down the outcome at each time point.

To account for the instructional sequence that takes place over multiple years and in multiple settings, Hong and Raudenbush (2006) proposed a method of causal inference with multilevel non-experimental data. The authors incorporated treatment weighting into a four-way hierarchical linear model with student i nested in the classroom j school k across t time points. The model was estimated using the

maximum pseudolikelihood function. Under the assumption of strong ignorability, the treatment given in a year is independent of all the future outcomes given the past outcomes and observables and there is no need to assume constant treatment effects (Hong & Raudenbush, 2006). The weight function is defined as

$$\omega_{itk} = \frac{h(z_{1ik})}{h(z_{1ik}|x_{1ik})} \times \frac{h(z_{2ik}|z_{1ik})}{h(z_{2ik}|z_{1ik},x_{1ik},x_{2ik},Y_{1ik})} \times \ldots \times \frac{h(z_{tik}|z_{1ik},\ldots,z_{t-1ik})}{h(z_{tik}|z_{1ik},\ldots,z_{t-1ik},x_{1ik},\ldots,x_{tik},Y_{1ik},\ldots,Y_{t-1ik})},$$

where h is the conditional likelihood of receiving either treatment or control.

3 Discussion

Education is an open and recursive system that involves a mutual and symbolically mediated relationship between students and teachers (Bieasta, 2007). We need causal inference models to consider this bi-directional and recursive relationship. As some instructional regimes require multiple years of instruction in social contexts, we need more methodological research with RCTs to examine the dynamic intervention effect with latent traits (e.g., incorporating proportion weighting to structural equation modeling). The intent-to-treat analysis is another research direction to study the effect of change of group membership on final outcomes for participants who did not follow the initially assigned treatment level. Lastly, meticulously planned studies for valid measurement of outcomes from interventions should be encouraged.

As stated before, both Hong & Raudenbush (2006) and Raudenbush (2008) are excellent examples of reconsideration and reframing of assumptions. Some of the assumptions can still be challenging to satisfy in a realistic education setting. For example, it is unlikely that each classroom j will have a unique teacher t_j given the long-existing teacher shortage problem in the United States and worldwide (Ingersoll & Smith, 2003). In reality, a teacher t might be responsible for teaching the same subject in multiple classrooms in the same school, even if an instructional intervention is assigned on a classroom basis. That is not to say that, as educational researchers, we should refrain from conducting rigorously designed RCTs. There is simply the need to promote a deeper understanding of some of the canonical assumptions in RCTs and causal inference methods.

The paper aims to broaden the scope of what it constitutes as evidence-based practice in education. Evidence-based practice should not be viewed as synonymous with RCTs. SCDs, for example, should be used as an alternative or complementary methodology. Odom (2021) commented that 'single-case' could be a misnomer, as SCD can include one classroom, one school, or the whole community instead of just one participant. In Special Education, single-subject research can serve as the springboard from which researchers can identify promising practices of large-scale clinical trials (Sutherland et al., 2018). Special attention should be paid to the contextual factors and identifying moderators so we can have an idea for whom the intervention is the most effective. In addition, there are a variety of statistical methods available for SCDs. Alternatively, qualitative research trades generalizability for a more in-depth investigation on a local level (Blanchard et al., 2014). This goes without saying that institutional and funding support is needed for studies parallel to RCTs. Lastly, having process evaluation accompanied RCTs will push the field towards mixed-method research with guidelines for conducting and evaluating the research procedure and outcome (Hitchcock, Johnson, & Schoonenboom, 2018).

The context-bound nature applies to both SCDs and RCTs and conclusions made from a study should only be generalized to participants, classrooms, and schools of similar characteristics (Odom, 2021). Documentation of the process of participant selection, recruitment, and randomization plays a crucial role in determining the relevant context of generalization (Odom, 2021). Independent replications and aggregation of studies using the same or even different methodologies are needed to establish the external validity of the conclusion from an RCT study.

4 Conclusions

With the appropriate reframing of assumptions and process evaluation, there is still a promising future for RCTs in education. The reframing of assumptions and process evaluation fit the RCTs in an authentic

education context and examine intervention components that lead to desirable outcomes. An absence of positive effects associated with the intervention does not mean that a rigorously designed RCT is a failure. Consequently, the lack of positive significant findings should not necessarily lead to the conclusion that RCTs are not worth the investment. Ending with the conclusion, this response paper recommends the use of RCTs in education settings where the SUTVA assumption could be satisfied as well as allow room for more research methodologies to be included in the repertoire of evidence-based practice.

5 Declarations

5.1 Competing Interests

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

5.2 Publisher's Note

AIJR remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

How to Cite this Article

A. Y. Zhang, "From Clinical Trial to Education: Methodologies, Assumptions, and Directions", *Int. J. Methodol*., vol. 1, no. 1, pp. 2–10, Mar. 2022. https://doi.org/10.21467/ijm.1.1.4668

References

Barnes, C., Yoong, S. L., Nathan, N., Wolfenden, L., Wedesweiler, T., Kerr, J., ... & Grady, A. (2021). Feasibility of a Web-Based Implementation Intervention to Improve Child Dietary Intake in Early Childhood Education and Care: Pilot Randomized Controlled Trial. *Journal of medical Internet research, 23*(12), e25902.

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.

Blanchard, R. D., Artino Jr, A. R., & Visintainer, P. F. (2014). Applying clinical research skills to conduct education research: important recommendations for success.

Boylan, M., & Demack, S. (2018). Innovation, evaluation design and typologies of professional learning. *Educational Research, 60*(3), 336-356.

Byrne, B. M. (2013). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming.* Psychology Press.

Cochrane, A. L. (1972). Effectiveness and efficiency: random reflections on health services.

Cohen, E., Uleryk, E., Jasuja, M., & Parkin, P. C. (2007). An absence of pediatric randomized controlled trials in general medical journals, 1985–2004. *Journal of clinical epidemiology, 60*(2), 118-123.

Connolly, P., Biggart, A., Miller, S., O'Hare, L., & Thurston, A. (2017). *Using randomised controlled trials in education.* Sage.

Connolly, P., Keenan, C., & Urbanska, K. (2018). The trials of evidence-based practice in education: a systematic review of randomised controlled trials in education research 1980–2016. *Educational Research, 60*(3), 276-291.

Dawson, A., Yeomans, E., & Brown, E. R. (2018). Methodological challenges in education RCTs: reflections from England's Education Endowment Foundation. *Educational Research, 60*(3), 292-310.

Delavar, F., Pashaeypoor, S., & Negarandeh, R. (2020). The effects of self-management education tailored to health literacy on medication adherence and blood pressure control among elderly people with primary hypertension: A randomized controlled trial. *Patient education and counseling, 103*(2), 336-342.

Espelage, D. L., Rose, C. A., & Polanin, J. R. (2016). Social-Emotional Learning Program to Promote Prosocial and Academic Skills among Middle School Students with Disabilities. *Remedial and Special Education, 37*(6), 323-332. http://dx.doi.org/10.1177/0741932515627475

Grässel, S., &, D. (2020). Recent advances in the treatment of osteoarthritis. *F1000Research, 9*.

Hedges, L. V., & Schauer, J. (2018). Randomised trials in education in the USA. *Educational Research, 60*(3), 265-275.

Hitchcock, J. H., Johnson, R. B., & Schoonenboom, J. (2018). Idiographic and nomothetic causal inference in special education research and practice: Mixed methods perspectives. *Research in the Schools*, 25(2), 56-67.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association, 81*(396), 945-960.

Hong, G., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association, 101*(475), 901-910.

Ingersoll, R. M., & Smith, T. M. (2003). The wrong solution to the teacher shortage. *Educational leadership*, 60(8), 30-33.

Linnan, L., & Steckler, A. (2002). Process evaluation for public health interventions and research.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms. *National Center for Special Education Research.*

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned?. *Educational Researcher, 48*(3), 158-166.

Mandy, W., Murin, M., Baykaner, O., Staunton, S., Cobb, R., Hellriegel, J., Anderson, S., & Skuse, D. (2016). Easing the Transition to Secondary Education for Children with Autism Spectrum Disorder: An Evaluation of the Systemic Transition in Education Programme for Autism Spectrum Disorder (STEP-ASD). *Autism: The International Journal of Research and Practice, 20*(5), 580-590. http://dx.doi.org/10.1177/1362361315598892

McKnight, L., & Morgan, A. (2020). A broken paradigm? What education needs to learn from evidence-based medicine. *Journal of Education Policy, 35*(5), 648-664.

Odom, S. L. (2021). Education of Students with Disabilities, Science, and Randomized Controlled Trials. *Research and Practice for Persons with Severe Disabilities, 46*(3), 132-145.

O'Toole, T., Logemann, J. A., & Baum, H. M. (1998). Conducting clinical trials in the public schools. *Language, Speech, and Hearing Services in Schools, 29*(4), 257-262.

Parker, A. T. (2009). Orientation and Mobility with Persons Who Are Deaf-Blind: An Initial Examination of Single-Subject Design Research. *Journal of Visual Impairment & Blindness, 103*(6), 372-377. https://ezaccess.libraries.psu.edu/login?url=https://www.proquest.com/scholarly-journals/orientation-mobility-with-persons-who-are-deaf/docview/61869802/se-2?accountid=13158

Pring, R., & Thomas, G. (2004). *Evidence-based practice in education.* McGraw-Hill Education (UK).

Raudenbush, S. W. (2008). Advancing educational policy by advancing research on instruction. *American Educational Research Journal, 45*(1), 206-230.

Robins, J. M., Hernan, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology.

Rubin, D. B. (1986). Comment: Which ifs have causal answers. *Journal of the American statistical association, 81*(396), 961-962.

Ruble, L. A., McGrew, J. H., Toland, M. D., Dalrymple, N. J., & Jung, L. A. (2013). A randomized controlled trial of COMPASS web-based and face-to-face teacher coaching in autism. *Journal of Consulting and Clinical Psychology, 81*(3), 566-572. http://dx.doi.org/10.1037/a0032003

Salinas Cisneros, G., & Thein, S. L. (2020). Recent advances in the treatment of sickle cell disease. *Frontiers in Physiology, 11*, 435.

Schneider, M. (2019, January). *Opening plenary: IES director welcome (Paper presentation). IES principal investigators meeting, Washington, DC.*

Siddiqui, N., Gorard, S., & See, B. H. (2018). The importance of process evaluation for randomised control trials in education. *Educational Research, 60*(3), 357-370.

Styles, B., & Torgerson, C. (2018). Randomised controlled trials (RCTs) in education research–methodological debates, questions, challenges.

Sutherland, K. S., Lewis-Palmer, T., Stichter, J., & Morgan, P. L. (2008). Examining the influence of teacher behavior and classroom context on the behavioral and academic outcomes for students with emotional or behavioral disorders. *The journal of special education, 41*(4), 223-233.

Sutherland, K. S., Conroy, M. A., Algina, J., Ladwig, C., Jessee, G., & Gyure, M. (2018). Reducing child problem behaviors and improving teacher-child interactions and relationships: A randomized controlled trial of BEST in CLASS. *Early Childhood Research Quarterly, 42*, 31-43.

Torgerson, D. (2008). *Designing randomised trials in health, education and the social sciences: an introduction*. Springer.

Wehman, P., Schall, C. M., McDonough, J., Graham, C., Brooke, V., Riehle, J. E., Brooke, A., Ham, W., Lau, S., Allen, J., & Avellone, L. (2017). Effects of an Employer-Based Intervention on Employment Outcomes for Youth with Significant Support Needs Due to Autism. Autism: *The International Journal of Research and Practice, 21*(3), 276-290. http://dx.doi.org/10.1177/1362361316635826

Publish your research article in AIJR journals-

- Online Submission and Tracking
- Peer-Reviewed
- Rapid decision
- Immediate Publication after acceptance
- Articles freely available online
- Retain full copyright of your article.

Submit your article at journals.aijr.org

Publish your books with AIJR publisher-

- Publish with ISBN and DOI.
- Publish Thesis/Dissertation as Monograph.
- Publish Book Monograph.
- Publish Edited Volume/ Book.
- Publish Conference Proceedings
- Retain full copyright of your books.

Submit your manuscript at books.aijr.org