Brandon Helding

Associate Editor

*Current Issues in Education*

cie@asu.edu

<http://cie.asu.edu>

March 24, 2010

Dear Mr. Helding:

We appreciate the opportunity to revise and resubmit our manuscript, Does it Work? A Guide to Investigating the Efficacy of Interventions in Educational Research, to *Current Issues in Education.* We were pleased that the reviewers were quite positive about the article, and, at least in the case of Reviewer A, provided extensive detailed feedback to guide us in the revision process.

We have addressed the reviewers’ concerns in the attached manuscript. Below, we detail each of the changes we made, in response to each concern, via bulleted points.

We look forward to hearing back from you.

Best wishes,

Michele Gregoire Gill

Debbie Hahs-Vaughn

University of Central Florida

Below are the responses to reviewer comments, marked by bullet points.

Review Notes: 312‐901‐1‐RV

I really enjoyed reading this paper and feel that it is a useful introduction to intervention

research techniques. The only verbiage that I drastically object to is the word “assure” in the

Randomized Experiments section of the paper. Please rephrase this less strongly. The rest of

my review is to aid in the readability of the paper. My hope is that you will follow at least some

of my suggestions and that we will be able to see it in print soon.

* This change has been made to the following (p. 6-7):   
  “This is because when designed rigorously and systematically, random assignment creates groups that are similar, on average, and any differences between groups can be attributed to the intervention rather than other factors (Shadish et al., 2002; Torgerson & Torgerson, 2008).“

**Randomized Experiments for Intervention Research:**

Random assignment does not “assure” anything. It is random, therefore all of your academically

accelerated students could end up in one group and all of your struggling students could end up in the other. Imagine if you tried to use an intervention designed to foster gifted students with a group of students whose days are spent mainly in special educational classes. If you got any results at all they might not ever be able to be reproduced, especially if you assumed that both sets of students had approximately the same distribution of ability. The best random assignment can do for the groups to “tend to be similar”.

* This change has been made to the following (p. 6-7):   
  “This is because when designed rigorously and systematically, random assignment creates groups that are similar, on average, and any differences between groups can be attributed to the intervention rather than other factors (Shadish et al., 2002; Torgerson & Torgerson, 2008).“

**Introduction:**

Overall this is a good introduction.

I did find it written as if it was to be presented orally rather than as an article (i.e. “How is this all related to today’s presentation?”) which was a little disconcerting.

* This entire introduction has been rewritten to remove references to an oral presentation (pp. 3-6). Specific examples of the changes include:
  + “How is this all related to today’s presentation?” has been deleted (p. 5). Other sentences have been changed to reflect a written, rather than oral, presentation (e.g., It is our hope in this paper” as opposed to “It is our hope today.”)
  + P. 4-6: Removed references to “you” and “your;” replaced with “one” and “they” and “their.”

Some of extraneous words could be eliminated (“Well, ” in second paragraph), (“And” and “and they” in third paragraph), and (“today” in sixth paragraph).

* “Well” has been deleted from p. 4.
* On page 4, “and” and “and they” has been deleted and replaced with “resulting in superficial changed to the curriculum.”
* On page 5, “today” has been deleted.

Additionally, I was confused by “a hefty chores” in the first paragraph and wondered whether you meant that research was not generalizable to schools (from research labs to schools), across schools (from school to school), or both to and across schools in the seventh paragraph.

* On page 3, we changed “hefty chores” to “significant household chores.”
* On page 5, we address the issue of generalizability by rephrasing this sentence to read “Unfortunately, for many years, educational research on interventions was of poor quality and not generalizable to schools because of either its poor design or lack of ecological validity.”

**Types of Research Questions:**

Good – no major revisions to suggest. However, again you might want to tighten up your verbiage.

* Minor changes in language were made on pp. 7-9 to tighten up our verbiage, such as “Investigative questions concern why things are the way they are, such as why does a particular phenomenon happen.” Also, extraneous question marks were replaced with periods for ease of readability.

**Research Designs – Causal Questions in Intervention:**

**Introduction:**

Please revise your sentence that includes “how effective is X on Y” making “These are valid and

important….” a separate sentence.

* This change has been made (p. 6) as follows: “As these new elements are introduced in a classroom or implemented across grades or school-wide, the question of “how effective is X on Y” and more specifically “does X cause Y” often arises. These are valid and important questions to answer.”

You might consider rephrasing the implementation sentence as: “… teaching method increase student performance in mathematics more than the previous method.”

* This change has been made as follows (p. 6): “For example, does the implementation of a new teaching method increase student performance in mathematics more than the previous method?”

The paragraph would be better if it were cut in two, with the second paragraph starting “It is

important”.

* This paragraph is now two paragraphs.

Additionally the three parts of the definition of causality (which are key to your discussion)

could be highlighted by placing them in a numbered list rather than hidden within the paragraph.

* This change has been made (p. 6) as follows:

“It is important to define what causality is in the context of intervention research. A causal relationship is such that the following occurs:

1. the cause (e.g., new intervention or curriculum) occurs prior to the effect (e.g., student performance in mathematics);
2. there is a relationship between the cause and the effect; and
3. the cause is the only plausible explanation for the effect (Shadish et al., 2002).”

The last sentence in this paragraph does not need to refer to “For the purposes of this paper” simply state “The following types of research designs to address causal question will be reviewed: a)…”

* This change has been made (p. 6) as follows: “The following research designs will be reviewed: a) randomized experiment; b) quasi-experiment; c) regression discontinuity; d) propensity score analysis; e) correlational; and f) action research and design experiments.”

**Randomized Experiments for Intervention Research:**

First paragraph: I would not use “For the purposes of this paper” I would start out with “social

experiment” and “field experiment” and then define experiment as you do in your second sentence.

* This change has been made as follows (p. 6): “What is referred to as an experiment has been more specifically labeled a 'social experiment' (p. 546) by Cook and Shadish (1994) or a ‘field experiment’ by Kerlinger & Lee (2000, p. 581).”

In this sentence you talk about the “gold standard” is this in educational research, psychological

research, social studies research… I am positive it is not the “gold standard” in research in the physical sciences.

* Clarification has been made (p. 7): “Random experiments, more specifically termed randomized control trials (RCT) (Shadish) or more loosely “true experiment” (Wallen & Frankael, 2001, p. 279), are often known as the "gold standard" (Shadish et al., 2002, p. 13) in educational research.”

Second paragraph: Could there be additional analytical issues than homogeneity when looking

individual’s data within clusters?

* This information has been clarified as follows (p. 7): “If the interest is on studying the student, however, randomly assigning at the cluster level introduces analytical issues such as potential homogeneity that exists within the clusters. Although beyond the scope of this paper, statistical procedures such as multilevel modeling are commonly used to address issues related to cluster random assignment (Raudenbush & Bryk, 2002; Shadish et al., 2002).”

Last paragraph: I would start the first sentence as “Other commonly used experimental designs include:” and then exclude the word “design” from the rest of the sentence. This would shorten the sentence and make if flow more easily.

* This change has been made (p. 7): “Other commonly used experimental designs include: 1) pretest-posttest control group; 2) alternative treatments design with pretest; 3) multiple treatments and controls with pretest; 4) factorial; 5) longitudinal; and 6) crossover (Shadish et al., 2002).”

Also, I would tack this paragraph onto the end of the preceding paragraph.

* The last two paragraphs have been grouped together.

**Quasi‐Experiments in Intervention Research:**

First paragraph: Again “gold standard” for whom?

* Clarification has been made (p. 7) as follows: “Although RCTs may be the gold standard in educational research because they most clearly allow the inference of causality, there are many instances in which RCTs are not ethical nor feasible (Rutter, 2007). “

Additionally your last sentence in this paragraph is

awkward. What is more difficult? What needs more attention in designing the study?

* Clarification has been made (p. 7) as follows: “This does not mean that inferences of causality cannot be made from quasi-experiments; however, causality is more difficult and more attention must be paid in designing the study in order to reduce the likelihood of other explanations for what has occurred (beyond that of the intervention).. .”

Last paragraph: In the first sentence you introduce you list saying these are designs you don’t need to include the word “design” within the list.

* ‘Design’ has been removed as follows (p. 8): “There are a number of different types of quasi-experiments including designs that: 1) do not have control groups (e.g., one group pretest-posttest, repeated treatment); 2) have a control group but do not have a pretest (e.g., posttest-only with nonequivalent groups); and 3) have both a control group and a pretest (e.g., untreated control group with dependent pretest and posttest samples, untreated control group with dependent pretest and posttest samples using a double pretest, cohort control group with pretest from each cohort).”

Also, if you want to give an example of each type, you only need one. They are examples not an exhaustive list.

* This has been revised so that examples (rather than an exhaustive list) are provided (p. 8): “There are a number of different types of quasi-experiments including designs that: 1) do not have control groups (e.g., one group pretest-posttest, repeated treatment); 2) have a control group but do not have a pretest (e.g., posttest-only with nonequivalent groups); and 3) have both a control group and a pretest (e.g., untreated control group with dependent pretest and posttest samples, untreated control group with dependent pretest and posttest samples using a double pretest, cohort control group with pretest from each cohort).”

After the list you tell us that you are going to discuss two lesser used types of quasi‐experiments but you leave the reader wondering why out of all the various types, you chose these two that are not used very frequently to discuss.

* The following sentence has been added to clarify (p. 8): “These designs were selected because they have great applicability in educational research and, of all the quasi-experimental designs, are the closest kin to a randomized experiment.”

**Regression Discontinuity**

First paragraph: In the first sentence, I’m not sure what you mean by “or great application to” in the first sentence

* This sentence has been revised as follows (p. 8): “Although regression discontinuity was introduced in the late 1950s (Campbell, 1984), with the exception of application to evaluate Title I programs in the mid-1960s (Trochim, 1980), it has been used sparingly in educational research. “

and when you say “may increase the power of” in the second sentence what would you increase the power of?

* This sentence has been revised to clarify as follows (p. 8): “However, there are many instances in which regression discontinuity is a better option than a quasi-experiment or may increase the power of the test when combined with a randomized experiment (Shadish et al., 2002)…”

Also, I would chop the second sentence into at least two (maybe three) separate sentences to improve readability.

* This has been done.

Second paragraph: When you illustrate combining RD with a randomized control (not controlled) element I would phrase it as “With the FCAT reading proficiency example this combined process could…” You might want to consider rephrasing to ease reading.

* This change has been made (p. 9): “With the FCAT reading proficiency example this combined process can be illustrated as follows. Students scoring above proficient do not receive developmental reading. Students scoring below proficient are randomly assigned to one of two different types of developmental reading: one is a new, innovative developmental approach and the other is the usual developmental reading approach.”

Last paragraph: You jump from examples in the previous paragraph to numbers of studies in this one without transition. Perhaps starting the paragraph with “As stated previously, regression discontinuity designs have been used infrequently….”

* This has been done (p. 9): “As stated previously, regression discontinuity designs have been used infrequently in education and the social sciences.”

Additionally, instead of “Currently, only…” , I would have written “ Only three… were found by the authors that have…”

* This change has been made (p. 9): “Only three refereed studies were found by the authors that have applied regression discontinuity within the past ten years (Bryant, Bryant, Gersten, Scammacca, & Chavez, 2008; Cahan, Greenbaum, Artman, Deluya, & Gappel-Gilon, 2008; Gormley, Gayer, Phillips, & Dawson, 2005).“

Then the following sentence could read “Note that all three studies found have been published since 2004.”

* This change has been made (p. 9): “Note that all three studies found have been published since 2004. “

Also, this last sentence seems to contradict your first sentence that says they were used in the 50’s & 60’s. Please clarify.

* Clarification has been made (p. 9): “Only three refereed studies were found by the authors that have applied regression discontinuity within the past ten years (Bryant, Bryant, Gersten, Scammacca, & Chavez, 2008; Cahan, Greenbaum, Artman, Deluya, & Gappel-Gilon, 2008; Gormley, Gayer, Phillips, & Dawson, 2005). Note that all three studies found within the past ten years have been published since 2004.”

**Propensity Score Analysis:**

First paragraph: In the first sentence you say that propensity score analysis is “likewise” considered a robust design. I assume you are stating “like RD” PS is robust but you didn’t state that RD was robust within your description of it. You might want to add a statement about robustness there also.

* The following has been added (p. 8) to RD: “Regression discontinuity is a robust quasi-experimental design is an excellent approach to take in instances where it is not possible or not ethical to design a randomized controlled trial.”

In the second sentence you use the term data to define data. You might consider “Data is considered ‘Observational’ if it has been….”

* This sentence has been revised as follows (p. 9): “Data is considered observational if it has been collected on students that does not involve manipulation.”

In the last sentence, I would use the word “could” rather the “would” when talking about balancing the group out.

* This change has been made (p. 9): “Researchers that use observational data run the risk of biased data because differences between groups may be due to the treatment or to pre-existing differences between the groups (that could have been balanced out through the random assignment process in a true experiment) (Rosenbaum, 1986). “

Second paragraph: In the first sentence, I don’t think you need to list examples of medical research since that is not our field of study, we’ll take your word for it. Perhaps a statement like “A quick search of the medical research literature revealed four studies in 2001 alone.”

* There are actually quite a number of propensity score studies in the medical area. Listed here are only a few examples. No change has been made to this sentence.

Also in this paragraph, you talk about “traditional matching” and the limited number of covariates. Please give us more information about what traditional matching is and why it has limits on the number of covariates. Remember you are talking to a wide range of readers.

* The following has been added (p. 9-10): “Traditional matching is conducted based on direct matching of covariates whereas propensity score analysis matches on propensity score. In traditional matching, therefore, researchers are limited in the number of covariates on which to match and thus the design of the study is comprised.”

Additionally, I was surprised by the use of the term “mistaken identity”. If you mean that the results would show a false positive or a false negative which would lead the researcher to make the wrong conclusions then say so.

* This has been clarified (p. 10): “This bias can then lead to a case of mistaken identity (e.g., false positive or false negative) with the treatment effect (Rosenbaum, 1991).”

Third paragraph: The first and second sentences say the same thing.

* The second sentence has been removed.

I would start the second sentence with “More specifically, the ….”

* The sentence has been revised (p. 10): “More specifically, the propensity score is the estimated chance of receiving the treatment based on pretreatment covariates (Rubin, 1997).”

In the third sentence I would change the word “this” to “the”.

* This change has been made (p. 10). “The beauty in propensity score analysis is the one score is created from the information of the covariates that can be applied in the model rather than applying multiple covariates in the model--greatly simplifying the model (Rubin, 1997).”

The parsimony sentence is rather awkward, perhaps “Unlike other procedures this procedure does not hinge on parsimony…”

* This sentence has been revised as follows (p. 10): “Unlike other procedures, propensity score analysis does not hinge on parsimony.”

As for the limitations sentence, I believe what you are trying to say is “One limitation with the use of propensity score analysis is that the assurance that there will …. between groups is based solely on the use of observed covariates.”

* This sentence has been revised as follows (p. 10): “One limitation with the use of propensity score analysis is that the assurance that there will not be systematic differences between groups is based solely on the use of observed covariates.”

In the next sentence I would use the word “tend” rather than “will” and I would start the last sentence in this paragraph with “However, sensitivity…”

* These changes have been made (p. 10): “In comparison, random assignment provides the assurance that systematic group differences tend to not exist based on both observed and unobserved covariates. However, sensitivity analysis can be examined to determine the probability that relevant but unobserved covariates were excluded from the model (Rosenbaum, 1991).”

Last paragraph: Please do not use the word “though”. Please use the word “although”, which “though” is an abbreviation of. I was confused as to how one could frequently underutilize something and how could you tell? Additionally, if it is underutilized then how is it more common in education and social science research? Also, more common than where else?

* The last sentence has been removed. The final paragraph now reads (p. 10): “Although computing propensity score analysis is a multi-step process, it does not require special statistical software. Additionally, primers (Luellen, Shadish, & Clark, 2005) and how-to tutorials are available (Hahs-Vaughn & Onwuegbuzie, 2006).”

**Correlational Designs:**

One minor change to suggest – The sentence starting “Possibly the most known…” is a bit awkward.

Perhaps “Possibly the best known examples are the multitude of correlational …”

* This change has been made (p. 10): “Possibly the best known example are the multitude of correlational studies that examined smoking and lung cancer (Wallen & Frankael, 2001).“

**Action Research:**

First paragraph: The sentence starting “Teacher researchers….” might be improved as “Teachers

researchers indicate that through their investigations they learn…. “

* This sentence has been revised as follows (p. 11): “Teacher researchers indicate that through their investigations they learn about their students (Fecho, 2000), their schools (Herr, 1999), and themselves (Hankins, 1998).“

The only sentence that I would maintain out of the two paragraphs following the numbered list is the sentence starting “Although there are cited…” I would include the gist of this sentence within the “There are different schools…” paragraph. The other parts of the two paragraphs following the numbered list seem to be at odds with the brevity with which you describe the other designs.

* This change has been made. The only parts of the two paragraphs after the numbered list are the following sentence: “Although there are cited benefits of teacher research, there are also concerns with effectively supporting teacher researchers and assisting teachers to learn about teacher research (Radencich, 1998).”

**What is the role of these designs in intervention research?**

First paragraph: The first two sentences here are the only ones I would revise. Contrary to the first sentences statement, I think you do present a rather good basic primer of the designs so I would rewrite it as “The descriptions presented are not only meant to serve as a primer on the designs but also to …” The second sentence is rather awkward. Please consider revising it.

* These revisions have been made to the first two sentences of the paragraph (p. 12): “The descriptions presented are not only meant to serve as a primer on the designs but also to provide a framework so that discussion on how they can be applied in intervention research is better contextualized. When applied effectively in intervention research, randomized control trials (RCT) will provide the best evidence of causality of all research designs.”

**The role of non‐causal research designs …**

This is a rather good section but again I do have a few suggestions. In your third sentence starting “Beyond these types…”, I would change “valuable to examine *that* do not…” to “valuable to examine *but* do not…” In the second to last sentence I would use the word “and” between cause and effect rather than the “ / “ for clarity. I would start the last sentence with “Additionally,” rather than “And”.

* Page 22: Changed “valuable to examine *that* do not…” to “valuable to examine *but* do not…”
* On page 23: Added the word “and” between cause and effect rather than the “ / .”
* Started the last sentence with “Additionally,” rather than “And”.

**Flowchart for determining how to ….**

Third paragraph: Consists of one sentence. I would and the word “a” before well‐designed and change the word “studies” to “study” in the first part of the sentence. I would then start a new sentence after “linear modeling)…” with “These analyses should control for pre‐existing…”

* Both changes made on page 25.

Fourth paragraph: I would add the word “be” before the word “limited” in the second sentence.

* Done, see page 25.

Fifth paragraph: Contains multiple run‐on sentences. In your first sentence, I would re‐write as “… of the designs, such as propensity score analysis, in their research. Figure 1, however, serves…”

* Done, see page 26.

The sentence starting “Applying randomized controlled…” I would end after the word “adopt”, delete the word “because” and start a new sentence with the word “Determining”.

* Done, see page 26.

I would also add the words “it is” before “also important”, rephrasing as “but it is also important for …”

* Done, see page 26.

Finally, in the last sentence I would re‐phase as “In the next section, we present an example of a study design for testing whether a particular…”

* Done, see page 26.

**Example study:**

This is a very good section. The one addition I would make is to the second paragraph. I would reiterate the three elements in the last sentence “In combination, these three elements (colleague observation, teaching log and lesson plans) will provide some…”

* Done, see page 27.

**Conclusion**:

There are some tense mismatch in the first paragraph in the sentence “The role of these designs…” should have “were” instead of “are” and “was” instead of “is”. Additionally the word also in the last sentence isn’t necessary.

* All recommended word changes made, see pp. 28-29.

In the last paragraph, you mention the “results” but you did not run any analysis or present any results. Perhaps simply “This paper was designed to serve dual purposes:…”

* Done, see p. 29.

**Response to Reviewer B**

Subject: Does it Work? A Guide to Investigating the Efficacy of Interventions in Educational Research

1. The purpose of study: The paper addresses a significant subject that would certainly make an appeal on the researchers and in particular the educational researchers. At the same time, there appears that the educational researches may not get a clarity as to why should they include this research method. Precisely, why would the educational researchers be motivated to apply this

intervention method? Though, what and how are dealt as effectively as they are required to be. Therefore, a little focus on this prime question would bring out the actual merit of this paper.

* The entire focus of our introduction is on the need for high-quality intervention research in education. We clarified this purpose further by adding the following sentence to page 4: “It would save much time and resources if teachers knew that what they were being told to implement in their classrooms really and truly works.”

2. Recommendation regarding the status of paper: Accept with revisions: The paper is recommended for publication provided the following inclusions/changes though minor but significant, are made.

A. Required Changes:

i.) A clearer approach to the target audience as to why should they take up this innovative method of research in education, may have to be dealt. Precisely, what would be the benefits in terms of outcome in the research, if this method is followed by the researchers, needs address?

* We address this concern on page 26: “Determining the efficacy of an intervention on a local scale is not only practically useful—it can help one’s day to day teaching—but it is also important for the broader research community in that it may reveal conditions under which the treatment X works or does not work as well.”

ii.) The prime objective of this paper appears to be bringing out the significance, need, scope and above all proving the merits of intervention method over the other methods but in the developmental process of this paper, a lot of analysis and methods used in other methods of research has been given more than required emphasis, which can be avoided.

* The purpose of the paper is to review multiple methods for conducting high quality research concerned with causal questions, particularly when straightforward experimental research cannot be conducted.

3. Feedback for the author/s

The paper has many merits such as cohesiveness, innovation, value in terms of rarity and its aptness but it needs a vertical development in which the area of its main objective occupies to be the main theme throughout, laying more and more focus on the selected subject with of course the other methods which have dealt well but not placed well. The paper deserves high appreciation because it addresses a subject that may bring in more openness and variety for conducting educational research.

* We clarified much of the language of the paper (see our response to Reviewer A), which should assist in streamlining the focus of our review to provide better “vertical development” of our ideas.