academicJournals

Vol. 10(2), pp. 111-118, 23 January, 2015
DOI: 10.5897/ERR2014.1835
Article Number: 01AA00749743
ISSN 1990-3839

http://www.academicjournals.org/ERR

Educational Research and Reviews

Full Length Research Paper

Internal validity: A must in research designs

Cahit Kaya

Department of Psychological Counseling and Guidence, University of Wisconsin- Madison, United State.

Received 2 May, 2014; Accepted 31 December, 2014

In experimental research, internal validity refers to what extent researchers can conclude that changes in the dependent variable (that is, outcome) are caused by manipulations to the independent variable. This causal inference permits researchers to meaningfully interpret research results. This article discusses internal validity threats in social and educational research using examples from the contemporary literature, and research designs in terms of their ability to control internal validity threats. An Eric and psychINFO search was performed to the current review of internal validity. The review indicated that appropriate research designs that control possible extraneous variables are needed to be able to meaningfully interpret research results in education and social sciences. Although, pretest-posttest experimental-control group design controls most of the internal validity threats, the most appropriate research design would vary based on the research questions or goals.

Key words: Internal validity, education, confounding variables, research design.

INTRODUCTION

One of the fundamental purposes of educational research is to determine possible relationships between variables. Experimental designs provide the best possible mechanism to determine whether there is a causal relationship between independent and dependent variables by applying an intervention to one group of research participants (that is, experimental group) while withholding it from another group (that is, control group). Subsequently, performances of the both groups on an outcome variable are compared to determine the possible effect of the intervention (Cook and Rumrill, 2005).

For example, Swiderski and Amadio (2013) examined the effectiveness of popular television clips as exemplars of Piagetian concepts compared to verbal descriptions of the same exemplars with a sample of college students. They concluded that an advantage in learning the concept of conservation at follow-up stage for students exposed to the popular television exemplars. However, there might be many other factors that can possibly influence the learning outcomes of the students. As a result, the researchers may not have strong confidence in reporting that the research results were precise.

The extent researchers can conclude that changes in the dependent variable (that is, outcome) are caused by manipulations in the independent variable is called internal validity. Cook and Campbell (1979) described internal validity as a phenomenon with which researchers infer that relationships between independent and dependent variables are not random but casual. In other words, interval validity of a study is a process to make sure changes in the dependent variable are due to the independent variable, not other confounding variables.

E-mail: ckaya@wisc.edu.

Internal validity is the sine qua non and conducting experiments with a strong internal validity is, therefore, ideal (Campbell and Stanley, 1966).

In order to establish strong internal validity, researchers need to, as much as possible, minimize confounding variables which are undesirable variables that influence the relationship between independent and dependent variables (Cook and Rumrill, 2005). Confounding variables might vary from one study to another depending upon experimental conditions. Examples of confounding variables include but not limited to changes in the history of participants, familiarity with experimenter and study conditions etc. One prevailing way to control confounding variables is employing research designs that provide strong internal validity and are compatible with research conditions (Campbell and Stanley, 1966).

Although it is relatively easier to establish internal validity in hard sciences, due to the complexity of human behaviors, it requires much more effort to do that in social and educational sciences. Social scientists cannot readily claim that treatment is the only cause of change in behavior. In order to have a strong internal validity and provide more accurate results, unlike in hard sciences, social scientists need to override other possible explanations of changes in behavior, such as developmental variations in participants, and changes in environmental conditions (Cook and Shadish, 1994).

The study examines internal validity threats observed in social and educational research illustrated by examples from contemporary literature. The study also discuss key features of research designs that help to control internal validity threats, and thereby enable researchers to interpret research results meaningfully.

REVIEW

Threats to internal validity

Campbell and Stanley (1966) in their classical book of experimental and quasi-experimental designs pointed out eight different classes of extraneous variables. They indicated if not controlled, those extraneous variables might have the confounding effect on the dependent variable. Those threats are explained in the following.

History

History threat refers to any events that happened between first and second measurement that influence outcome variable in addition to experimental variable. Those events can be anything that affects the results. Although most of those events might closely be related to participants` and studies` characteristics; changes in public policy (for example, changes in public school policies) and weather conditions (for example, improving mood because people are outside more) could also have history effect (Cook and Rumill, 2005). To exemplify history threat, a student with emotional disturbance might start displaying more socially appropriate behaviors after an intervention program, however the student`s appropriate behaviors might not solely be due to the intervention but also as a result of the students` parents teaching the student appropriate behaviors during the intervention.

Maturation

It refers to processes that changes in respondents` actions are simply due to the passage of time (not due to a specific event) in addition to the effect of the experimental variable. Changes in respondent`s actions might be as functions of biological or psychological changes that occur to participants, including growing older, growing hungrier, growing more tired and the like. Maturation could be a problem especially in longitudinal studies where the impact of an intervention is measured over a long period of time (Onwuegbuize, 2003). To exemplify it, a teacher who teaches a class at the end of the day might not obtain educational improvements as much as other teachers and assume that his/her teaching method is not affective, however, the reason he/she does not obtain better educational outcomes might be that the students feel exhausted at the end of the day.

Testing effect

It indicates the influence of taking tests one after another. It is also called practice effect. With repeated test taking, test takers will obtain better outcomes, not because of interventions but because they will become more familiar with test questions and formats. Although using different tests might reduce the possibility of getting higher results, some improvements might occur anyway due to participants` increasing familiarity with testing procedure (Rogers and Holloway, 1990). For example, if a teacher wants to measure the progress of a student with learning disability and give similar standardized test over and over, then the student might pick up the correct answer and have high scores as a result of his/her familiarity with the tests.

Instrumentation

It refers to having drifts in the calibration of a measurement tool, changes in scorers and observers that may impact results of the measurements. Instrument effect is most commonly seen when different instruments are used in pretests and posttests or a measurement tool is administrated by different researchers. In such situations, changes in outcomes might be attributed to inconsistencies in instrumentation (Campbell and Stanley, 1966). To exemplify instrumentation threat, some teachers tend to give higher scores on exams than other teachers. If an exam is graded by two teachers, the students` results on exams are prone to be influenced by the teachers` tendencies in scoring.

Statistical regression

Participants with extreme scores at first measurement tend to move toward mean at the second measurement which confounds the results. This situation can be explained by two factors:

1. A measurement always consists of true score and a chance of events, and
2. In the distribution of scores, there are always more scores combined around mean then around sides of extreme scores (Cook and Rumrill, 2005).

Deleting participants (in both the experimental and control groups) with extreme scores can reduce the effect of regression to mean. For example, if a researcher wants to decrease aggression among students and apply an intervention only to students who received extreme aggression scores in pretest; it is likely that regression to mean occurs in the study. It is expected that the students` extreme scores will move toward mean at the posttest.

Selection

It refers to biases in the selection of participants for experimental and control groups. It is most commonly experienced at the data collection stage where already-formed groups are compared. Group differences as a result of selection bias can occur in terms of cognitive, affective, personality and demographic variables. Selection bias may always occur, however, the lesser this bias, the higher internal validity is achieved (Onwuegbuize, 2003). For example, if a researcher wants to investigate how an educational intervention can increase academic achievement and only chose participants from high achieved schools; the intervention would most likely produce positive results. However, the reason for the positive results might be that the participants from high achieved schools are more motivated to learn than the other school students.

Experimental mortality

It refers to the loss of participants in different rates in experimental and control groups. The loss can be due to death or drop out from studies. A high mortality rate can cause selection biases if mortality rates are differentiated within experimental and control groups. Results of studies will be confounded if mortality has a specific relationships with particular characteristics of either experimental or control group participants (Cook and Rumrill, 2005). For example, if a researcher wants to investigate effects of a class intervention on students` academic success and require attendance only for the experimental group, in meantime, the control group might shrink and only high motivated students might remain in the control group. In this case, outcomes of the control group will be influenced by the remaining control group participants` characteristics.

Selection-maturation interaction

It basically refers to the selection of comparison groups interacting with history, maturation and mortality threats influencing the results of studies and causing false interpretation of that treatment caused the effect. Selection history interaction occurs if individuals in the groups experience different history events; selection maturity interaction occurs if when one group has higher maturation rate than another group and selection mortality interaction occur if one group has higher attrition rate than another group (Onwuegbuize, 2003). Kirk (1995) in identifying internal validity threats focused on more social aspects of experiments. He, in addition to Stanley and Campbell (1966), proposed three other internal validity threats:

1. Demand characteristics.
2. Participant predisposition affect, and
3. Experimenter expectancy effect. Those threats are explained at the following.

Demand characteristics

It refers to cues in experimental environment or procedure that lead participants to figure out the aim of experiment and respond it purposely either in the positive or negative way. Participants might get cues from rumors about an experiment, laboratory environment (for example, cameras), and communication occurs when meeting with experimenters. Demand characteristics cue participants what is expected and thus impact participants' behaviors.

Participant-predisposition effect

It refers to participants` history, personality, or predisposition to respond to experiment in a particular way. He identified two types of participants` affect. The first effect is called ‘cooperative participants affect' in which participants unconsciously or consciously provide data that support the experiment. The second effect is called ‘screw you affect' in which participants consciously or consciously provide data that do not support the experiment. The reasons participants act in that way might be resentment over participating in an experiment, or having bad experiences with previous experiments such as being deceived, or made to feel inadequate.

Experimenter-expectancy effect

It refers to experimenters acting in a way to convey how they expect participants to behave during experiments. Experimenters while communicating with participants may deliver a more subtle request or messages. For example, experimenters body language, the tone of voice and facial expression can cue researcher`s expectations, and thus affect participants` performance. He indicated experimenters` expectancy may also influence in a way how experimenters collect, record, analyze and interpret data. Although, small errors may always be accompanied by data collection and analyzes, more often occurrences of those errors may lead to significant changes in the direction of results.

Various ways to minimize threats to internal validity

In order to minimize threats to internal validity, researchers need to carefully design their studies in accordance with their research conditions. Although, it is challenging to eliminate extraneous variables, and employing strong research designs will help to minimize internal validity threats (Mitchell and Jolley, 2010). In the current review, two true experimental designs, one quasi-experimental design, and one pre-experimental design will be examined in terms of their power to minimize threats to internal validity. Although, quasi-experimental and pre-experimental designs are less used in educational research, they are examined to make cross-comparison between experimental designs. A summary of experimental designs and their ability to control internal validity threats are provided in Table 1.

True experimental designs

There are three types of true experimental design:

1. Pretest-posttest control group design
2. Posttests only control group design, and
3. The Solomon four-group designs.

These designs are mostly recommended in literature as they provide highest internal and external validity (Campbell and Stanley, 1966). The Solomon four-group design combines pretest-posttest and only post-test control group designs and thus requires relatively a high amount of time and resources. As a result, it has less applicability in social and educational research. Therefore, this design was not included in the current review.

The pretest-posttest control group design

This design is regarded as the gold standard among experimental designs. It controls for all of the aforementioned threats utilizing a control group, random selection and random assignment (Campbell and Stanley, 1966). In following, the way how these threats are controlled will be explained.

History is controlled insofar as it is expected in general events that occur between first and second measurement which would affect experimental and control group equally. However, this design cannot control intra-session history effect which is events occurring specific to (for example, obstreperous joke) experimental and control group session. It is recommended both experimental and control group sessions to be conducted simultaneously to reduce the possibility of intra-session history effect (Campbell, 1957).

Maturation and testing are controlled insofar as it is expected they should manifest in experimental and control groups equally. Both experimental and control group participants would in average have the equal maturity or familiarity with testing (Tucker-Drob, 2011).

Instrumentation is controlled as intra-session history is controlled. In a case of few observers and interviewers, where they cannot randomly be assigned to single experimental and control group session, to control instrumentation, each observer and interviewer need to be assigned both experimental and control groups sessions and they should be ignorant of what session is being conducted. The use of recordings of group interactions in pretest, posttest for experimental and control group sessions that could be reviewed by judges helps to control instrumentation effect as well (Campbell and Stanley, 1966).

Table 1. Experimental designs and their ability to control internal validity threats.

Variable	History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction of Selection and maturation etc.
True experimental design	+	+	+	+	+	+	+	+
Pretest-posttest control group design								
Posttest control group design	+	+	+	+	+	+	+	+
Quasi experimental design								
Time serious design	-	+	+	?	+	+	+	+
Pre-experimental design								
One-shot case study	-	-				_	_	
One group pretest posttest design	-	-	-	-	?	+	+	-
Static group comparison	+	?	+	+	+	-	-	-

Note. Adapted from Campbell and Stanley`s (1966) classical book of experimental and quasi-experimental designs. In the tables a minus indicates a definite weakness, a plus indicates that the factor is controlled, a question mark indicates a possible source of concern, and a blank indicates that the factor is not relevant. It is noted that the table may not represent the more complex and qualified presentation in the text. Readers should not depend upon the table to understand the complex designs and their ability to control internal validity threats. Readers are strongly encouraged to comprehend what each + and – means in experimental designs to be able to interpret the table.

Regression is controlled as far as mean differences are concerned. Randomly assigning participants from a pool of extreme pretest scores to experimental and control groups will result in both experimental and control group regressing at equal levels (Campbell, 1957; Campbell and Stanley, 1966).

Selection is ruled out as participants are randomly selected from the population and randomly assigned experimental and control group conditions (Walker, 2005). Random selection and assignment are aimed to achieve no difference hypothesis between experimental and control groups as much as possible. In accordance with sample in the statistic and larger sample size, the greater equality between groups is achieved (Campbell and Stanley, 1966).

Although hypothetically no difference is assumed between experimental and control groups, there occasionally would be a significant difference in between pretest scores of the experimental group. While simple and stratified randomization provides unbiased assignment of participants to groups, it is worth to note that it is less than a perfect of way promising initial equivalence of groups. However, it is the only and essential way of doing so (Campbell and Stanley, 1966).

Mortality is controlled insofar as it is expected that both experimental and control groups will be influenced by mortality, lost cases, and partial data equally. However, in some cases, mortality can provide a plausible explanation for differences between experimental and control groups. In cases such, researchers who eliminate participants who failed to show up for experimental sessions are biasing experimental group in the direction of conscientious and healthy. The suggested way of dealing with this issue is including pretest and posttest of scores of participants who failed to show up in data analysis. Although this procedure attenuates the apparent effect of the independent variable, it rules out sampling bias. Another way of partially dealing with it is comparing numbers and pretest scores of participants who were present at the pretest but not at the posttest with those who were present both at the pretest and posttest (Campbell, 1957; Campbell and Stanley, 1966).

The posttest control group design

In this design, participants are randomly selected from the population and randomly assigned to both control and experimental groups. Yet, there are no pretest measurements of participants. However, through random selection and random assignment, the equivalence of experimental and control group is assumed. In real life there are few situations where pretest occurs, therefore this design is considered more representative of real life settings (Mitchell and Jolley, 2010). This design is recommended when pretest-posttest control group design is not applicable. It is particularly suggested for educational settings where the introduction of an entirely new subject matter throughout the year is a common practice, as a result, it is hard to have continues pretest measurements before each subject is presented to compare pretest and posttest scores (Campbell and Stanley, 1966).

This design technically rules out history, maturation and testing effect most of the time. It is expected that general events that happen to experimental groups will also happen to control groups. Though, an intra-session history is still an intact threat. As for maturation, it is expected that both experimental and control groups participants have a similar manifestation of maturation and in usual, there is less time between presentation of the independent variable and posttest measurements. There is no pretest measurement so testing and regression to mean threats are ruled out automatically. Instrumentation is controlled as far as intra-session history is controlled. However, random assignment of observers and interviewers to experimental and control groups without them knowing which group is receiving intervention is recommended. Selection is ruled out since participants are randomly selected from the population and randomly assigned experimental and control group conditions (Campbell, 1957; Campbell and Stanley, 1966). As for mortality, although pretest-posttest control group design provides more opportunity to detect differential mortality, this design relatively controls for mortality effect by assuming that through randomization mortality rates would be equivalent for experimental and control groups (Jurs and Glass, 1971).

This design is more convenient than pretest-posttest control group design as a result of not including pretest measurements (Mitchell and Jolley, 2010). In educational settings where conducting pretests seem awkward and influence results of the independent variable, application of this design is suggested (Campbell and Stanley, 1966).

Quasi-experimental designs

Quasi-experimental designs have a similar purpose of finding a causal relationship between dependent and independent variables. However, by definition, quasi-experimental designs are the lack of random assignment. Assignment to conditions is achieved by a variety of ways such as researcher selection (Levy et al., 2011). Quasi-experimental designs include time-serious design, the equivalent time-samples design, the equivalent materials design, and non-equivalent control group design (Morgan et al., 2000). However, due to space constraints only time serious design which relatively resembles multiple baseline single-subject research design will be explained.

Time serious design

This design measures several waves of observation before and after the introduction of the independent (treatment) variable. The problem of internal validity in this design is expressed as possible alternate explanations of the shift in the time series other than the effect of the independent variable. The strengths of this design are not needing a control group and having timely measurements before and after the introduction of independent variable (Campbell and Stanley, 1966).

Campbell and Stanley (1966) indicated the most plausible threat to internal validity of this design would be the failure to control history. The rival hypothesis would be that not the dependent variable but some other events simultaneously caused the shift. For example, a research investigating the effect of a documentary film on students` optimism might fail to provide a clear cut control on history. Because the students might daily be exposed to many other potential relevant sources. Possible history threats in addition to simultaneous events (for example, effects of weather and effect of seasons in case of extended studies) may influence the results. The effect of history threat can be minimized via experimental isolation in which possible other variables causing the shift are controlled.

Campbell and Stanley (1966) indicated maturation threat will be ruled out because it is expected no abrupt changes will be observed during pretest measurements (before the introduction of independent variable) and changes are expected to occur after the independent variable is presented. Both instrumentation and testing effect are eliminated because there would not be a specific rationale that an error particularly occurs after the presentation of the independent variable. Regression would not occur because it would be implausible that an affect at far along posttest measurements is greater than the effects at earlier pretest measurements. Selection is ruled out because same participants are involved in all measurements. Since the data is collected on individual group members, there would not be specific mortality effect.

Pre-experimental designs

Pre-experimental designs only follow basic experimental steps and do not include control groups. Therefore, those studies have almost a total absence of control, and no scientific value (Campbell and Stanley, 1966). These designs are recommended when none of the true experimental and quasi-experimental studies can be conducted. There are three types of pre-experimental designs:

1. One-shot case study
2. One group pretest-posttest and
3. The static-group comparison designs (Campbell, 1957).

In pre-experimental studies, no internal validity is achieved. However, one group pre-test post-test study includes only one group, differential rates of mortality and selection bias are ruled out automatically (Campbell, 1957; Campbell and Stanley, 1966).

METHODOLOGY

The current study searched Eric and Pscyhinfo databases to retrieve articles regarding internal validity. Key words ‘internal validity threats’ and ‘education’ were searched utilizing Eric and PscyhINFO. Two criteria were identified to identify papers in this review:

1. The article was in the field of social and educational sciences.
2. The article had a focus on the internal validity or internal validity threats.

The results indicated 84 articles. The articles were studied to prepare the current review of the internal validity and internal validity threats in educational and social sciences. Examination of the research articles pointed out that most of them referred to Campbell and Stanley (1966) classical book of experimental and quasi-experimental designs as the original source. Therefore, Campbell and Stanley (1966) book was used as a main source in this study as well.

DISCUSSION

Internal validity is an internal part of experimental studies to ensure that the relationships between independent variable are casual and not confounded with extraneous (confounding) variables. Internal validity is an important issue that needs to be addressed particularly in social sciences due to the involvement of human subjects in experimental studies. Internal validity within experimental studies can be increased by minimizing influences of confounding variables on the dependent variable. Confounding variables might vary from one study to another and can influence researches in various ways. Researchers need to carefully design their research to have the highest control over confounding variables.

Researchers identified several lines of threats to internal validity. Campbell and Stanley (1966) identified eight threats to internal validity which are most commonly referred in the literature. Kirk (1994) added three other internal validity threats that were typically linked to participant`s, and researchers` characteristics, and experimental conditions. If not controlled, all of the internal validity threats could have a dramatic effect on research results and attenuate accuracy and validity of research results. When true, quasi and pre-experimental designs were examined, the most challenging threat to internal validity was history threat. Although history threats could be controlled by true experimental designs, intra-session history threat was still a hurdle for these designs. However, several ways of minimizing intra-session history exist.

True experimental designs were most recommended designs due to their ability to control most of the internal validity threats by having the random selection, random assignment and a control group. Through random selection and assignment, it was assumed that changes in participant`s, researcher`s and experimental condition will have an equal level of influence on experimental and control groups. In other words, changes in one group will be overridden with similar changes in another group. Although all of the true experimental designs minimize internal validity threats, thus achieve strong internal validity, the posttest only control group design was recommended in educational settings as it does not require pretest measurements.

In comparison to true experimental designs, quasi-experimental designs, and pre-experimental designs provide either lesser or no control on internal validity threats. When applicability of true experimental designs is limited, employment of those designs can be considered as an option to control internal validity threats. Quasi-experimental designs employ same strategies as true experimental designs; therefore can provide relatively strong internal validity. However, due to lack of random assignment, quasi-experimental designs are susceptible to many internal validity threats. Pre-experimental designs are the lack of internal validity. When none of other research designs is an option, these designs can be considered as an option.

Conclusions

Internal validity is a must for researchers to conclude that the relationships between independent and dependent variables are not random but casual. Researchers in education and other social sciences must understand what internal validity is and how internal validity is achieved. Without having a strong internal validity in their studies, researchers cannot confidently claim that their results are accurate.

There are different types of internal validity threats. Depending upon research design and conditions, internal validity threats may vary. However, all of the internal validity threats may have a strong influence on research results. Researchers need to contemplate on the aforementioned or any other internal validity threats that may occur in their studies before initiating a study. Using a strong research design and modifying their studies in accordance with their research conditions, researchers may be able to control internal validity threats that may affect the accuracy of their results.

The most common internal validity threat among various research designs is history threat. Although true experimental designs can control history threat, intra-session history threat is still a hurdle for these designs. Conducting experimental and control group sessions at the same time and randomly assigning researchers to experimental and control groups in a way that the researchers would not know which experimental group they were assigned is a way to control intra-session history threat.

Considering the power of research designs in eliminating internal validity threats, researchers should employ research designs that are most appropriate to their research conditions. True experimental designs provide the most control over internal validity threats and should be preferred by the researchers. However, in the case of true experimental designs cannot be utilized, quasi and pre-experimental designs can be considered as a second and third option respectively.

LIMITATIONS AND FUTURE STUDIES

This study is limited in several ways. First, internal validity is a very large theme and the current research only covered a small part of it. This paper had limited discussion about the definition of internal validity, internal validity threats and experimental designs in regard to their ability to control the internal validity threats. A more extensive research will provide a more comprehensive picture of internal validity within experimental studies. Secondly, this research focused on internal validity threats that are most common in the literature. Although those internal validity threats are important to control, there might other internal validity threats that need to be controlled. Further investigation of various internal validity threats will provide more information in regard to internal validity threats.

Thirdly, mostly book chapters explain the internal validity of research studies. Research articles that explain internal validity threats are very limited. Moreover, the literature was the lack of providing examples on how internal validity threats can be controlled within various research studies. Although this paper provides an overview of internal validity threats, it does not exemplify the actual application of how to control internal validity threats within particular research studies. Future studies may provide exemplify how to control internal validity threats within particular research studies.

CONFLICT OF INTERESTS

The author has not declared any conflict of interests.

REFERENCES

Campbell DT (1957). Factors relevant to the validity of experiments in social settings. Psychological Bulletin 54(4):297-312. doi:10.1037/h0040950.

Campbell DT, Stanley JC (1966). Experimental and quasi-experimental designs for research. Boston: Houghton Mifflin Company.

Cook L, Rumrill PR (2005). Internal validity in rehabilitation research. Work: J. Prevention, Assessment Rehabilitation 25(3):279-283.

Cook TD, Campbell DT (1979). Quasi-experimentation: Design and analysis issues for field settings. Chicago, IL: Rand McNally/

Cook TD, Shadish W (1994). Social experiments: some developments over the past fifteen years. Annual Review of Psychology 45545-580. doi:10.1097/00004583-200008000-00024.

Jurs SG, Glass GV (1971). Effect of experimental mortality on the internal and external validity of the randomized comparative experiment. Journal of Experimental Education 4062-4066.

Levy Y, Ellis TJ, Cohen E (2011). A Guide for Novice Researchers on Experimental and Quasi-Experimental Studies in Information Systems Research. Interdisciplinary Journal of Information, Knowledge Management 6151-6161.

Mitchell ML, Jolley JM (2010).Research design explained: Instructor's edition (7th ed.). Belmont, CA: Wadsworth/Cengage Learning.

Morgan GA, Gliner JA, Harmon RJ (2000).Randomized experimental designs. J. Am. Acad. Child Adolescent Psychiatry 39(8):1062-1063.

Rogers JC, Holloway RL (1990). Assessing threats to the validity of experimental and observational designs. Family Practice Research Journal 10(2):81-95.

Swiderski DJ, Amadio DM (2013). Teaching method and effect on learning Piagetian concepts. Psychology Learning and Teaching 12(1):55-63. http://dx.doi.org/10.2304/plat.2013.12.1.55

Tucker-Drob EM (2011). Individual differences methods for randomized experiments. Psychological Methods 16(3):298-318. doi:10.1037/a0023349.

Walker W (2005). The strengths and weaknesses of research designs involving quantitative measures. Journal of Research in Nursing 10(5):571-582. doi:10.1177/136140960501000505.