

5-22-2023

A Pilot Study of Development and Validation of Treatment Fidelity Checklist for Intervention Research in K-12 School Physical Education

Weidong Li

The Ohio State University, Li.832@osu.edu

Ping Xiang

Texas A & M University, ping-xiang@tamu.edu

Follow this and additional works at: <https://scholarworks.boisestate.edu/ijpah>



Part of the [Health and Physical Education Commons](#)

Recommended Citation

Li, Weidong and Xiang, Ping (2023) "A Pilot Study of Development and Validation of Treatment Fidelity Checklist for Intervention Research in K-12 School Physical Education," *International Journal of Physical Activity and Health*: Vol. 2: Iss. 3, Article 3.

DOI: <https://doi.org/10.18122/ijpah.020303.boisestate>

Available at: <https://scholarworks.boisestate.edu/ijpah/vol2/iss3/3>

A Pilot Study of Development and Validation of Treatment Fidelity Checklist for Intervention Research in K-12 School Physical Education

Abstract

No conceptual framework and valid and reliable measurement of treatment fidelity are available for intervention research in K-12 school physical education. The purposes of this study were to propose a conceptual framework and then develop and validate a checklist to assess treatment fidelity in K-12 school physical education. A literature search was conducted on journal article titles using key words through EBSCO Discovery Service. A conceptual framework and Treatment Fidelity Checklist were developed. The conceptual framework and Treatment Fidelity Checklist were validated by four experts and an intercoder reliability coefficient. All experts validated the conceptual framework and Treatment Fidelity Checklist. The overall IOA was 88%. The Treatment Fidelity Checklist is valid and reliable. The conceptual framework and Checklist of Treatment Fidelity provide a very useful document to guide future intervention research in physical education and what to be reported in publications.

International Journal of Physical Activity

**A Pilot Study of Development and Validation of Treatment Fidelity Checklist for
Intervention Research in K-12 School Physical Education**

Abstract

No conceptual framework and valid and reliable measurement of treatment fidelity are available for intervention research in K-12 school physical education. The purposes of this study were to propose a conceptual framework and then develop and validate a checklist to assess treatment fidelity in K-12 school physical education. A literature search was conducted on journal article titles using key words through EBSCO Discovery Service. A conceptual framework and Treatment Fidelity Checklist were developed. The conceptual framework and Treatment Fidelity Checklist were validated by four experts and an intercoder reliability coefficient. All experts validated the conceptual framework and Treatment Fidelity Checklist. The overall IOA was 88%. The Treatment Fidelity Checklist is valid and reliable. The conceptual framework and Checklist of Treatment Fidelity provide a very useful document to guide future intervention research in physical education and what to be reported in publications.

Keywords: fidelity of implementation, experimental research, treatment fidelity, treatment integrity, treatment differentiation

A Pilot Study of Development and Validation of Treatment Fidelity Checklist for Intervention Research in K-12 School Physical Education

Numerous factors determine the scientific rigor of intervention studies in k-12 school physical education, including research design, sampling, teachers' training, and treatment fidelity. As one of many critical factors, treatment fidelity mediates the effect of program interventions on program outcomes (Borrelli, 2011). Treatment fidelity is defined as "the ongoing assessment, monitoring, and enhancement of the reliability and internal validity of a study" (Borrelli, 2011, p. 552). It consists of two general components: a) treatment integrity, the degree to which interventions are implemented as planned or intended, and b) treatment differentiation, the degree to which conditions differ in critical dimensions (Borrelli, 2011). The importance of treatment fidelity in program interventions has been widely discussed and various conceptual frameworks and measurements have been developed to study treatment fidelity in both public health (Borrelli, 2011) and general education (Dane & Schneider, 1998; Durlak & Dupre, 2008; Dusenbury et al., 2003; Mihalic, 2004; O'Donnell, 2008).

In public health, Borrelli (2011) proposed a conceptual framework of treatment fidelity and developed a checklist to assess treatment fidelity. This conceptual framework is composed of five components: (a) treatment design, involving an operationalization of treatment components to be reflective of and aligned with the theoretical framework. The hypothesized intervention components are explicitly described in the intervention protocol, and in the training of providers and follow-up supervision of providers; (b) training providers, involving "standardizing training between providers, ensuring that providers are trained to criterion, and monitoring and maintaining provider skills over time" (Borrelli, 2011, p. 4); (c) treatment delivery, consisting of treatment differentiation, treatment competency, and treatment adherence. Treatment differentiation addresses whether the providers only deliver the treatment being targeted, not other treatments. Treatment competency involves the maintenance of skills that providers have learned during training. Treatment adherence is to ensure that the treatment components are delivered as intended; (d) treatment receipt, defined as whether participants demonstrate understanding and knowledge of what are delivered in the treatment and can use the learned skills; (e) treatment enactment, involving assessing, monitoring, and improving participants' ability to perform behavioral skills and cognitive strategies learned in the treatment in relevant real-life settings. Based on this conceptual framework, Borrelli (2011) developed a checklist for treatment fidelity, which consists of 29 items. The details of the checklist are presented in Table 1.

Researchers have examined treatment fidelity narrowly by mainly focusing on fidelity of implementation (FOI) of interventions rather than broadly by including treatment design and training providers. In some cases, this may be due to uniqueness of intervention research in educational settings. FOI refers to "the extent to which delivery of an intervention adheres to the protocol or program model originally developed" (Mowbray et al., 2003, p. 315). Researchers have proposed a five-element conceptual framework to assess FOI (Dane & Schneider, 1998; Durlak & Dupre, 2008; Dusenbury et al., 2003; Mihalic, 2004; O'Donnell, 2008). These consist of: (a) Adherence, referring to "whether the components of the intervention are being delivered as designed"; (b) Duration or Intervention dose, defined as "the number, length, or frequency of sessions implemented"; (c) Quality of program delivery, meaning "the manner in

Table 1
Twenty-nine Item Treatment Fidelity Checklist

Categories	Items
Treatment Design	<p>The authors shall provide information about treatment dose in the intervention condition, including length of contact, number of contacts, content of treatment, and duration of contact over time.</p> <p>The authors shall provide information about treatment dose in the comparison condition, including length of contact, number of contacts, content of treatment, duration of contact over time, and methods to ensure an equivalent dose between conditions and for participants within conditions.</p> <p>The provider’s credentials that are needed for treatment implementation shall be specified.</p> <p>The authors shall clearly articulate the theoretical model upon which the intervention is based by specifying what are the active ingredients and whether they are incorporated into the intervention, using experts to determine whether the intervention protocol is aligned with the underlying theoretical model, and having a plan to ensure that the measures are reflective of the hypothesized theoretical constructs or underlying mechanisms.</p> <p>The authors shall identify the potential confounders at the end of the study and have a plan to deal with possible setbacks during treatment implementation.</p> <p>If there are more than one intervention, the authors shall describe all the intervention conditions equally well.</p>
Training Providers	<p>The authors shall describe how providers will be trained through a training manual</p> <p>The authors shall standardize the provider’s training to eliminate the discrepancies in training.</p> <p>The authors shall provide data about whether the provider acquires the skill sets that are needed to implement the treatment.</p> <p>The authors shall assess and monitor provider’s skill maintenance over time.</p> <p>The authors shall clearly articulate what characteristics are sought or avoided in a treatment provider.</p> <p>The authors shall assess whether there is a match between the treatment provider and the intervention or not during hire.</p>

International Journal of Physical Activity

Delivery of Treatment	<p>The training plan shall take trainees' education, experience, and learning styles into consideration.</p> <p>There is a method to ensure that the intervention content is implemented as intended.</p> <p>There is a method to ensure that the intervention dose is implemented as specified.</p> <p>The authors develop a mechanism to assess the adherence of the providers to the intervention plan.</p> <p>The authors assess nonspecific treatment effects during intervention.</p> <p>The authors use treatment manual.</p> <p>The authors have a plan to assess if the active ingredients were delivered.</p> <p>The authors have a plan to assess if prescribed components were delivered.</p> <p>The authors have a plan for how to prevent contaminations from happening between conditions.</p>
Receipt of Treatment	<p>There is a benchmark for treatment fidelity being specified (e.g., providers delivering 80% or above of components).</p> <p>The authors assess the extent to which participants demonstrate an understanding of the intervention.</p> <p>Strategies to improve participants' comprehension of the intervention are specified.</p> <p>The authors will assess participants' ability to perform the intervention skills during intervention.</p> <p>A strategy is used to increase participants' performance of intervention skills during intervention.</p>
Enactment of Treatment	<p>Multicultural factors such as use of language and values of the target group will be considered when the authors develop and deliver the intervention.</p> <p>The authors shall assess whether participants can perform the skills in settings where the intervention might be applied.</p> <p>The authors shall indicate whether a strategy is employed to assess participants' performance of the skills in settings where the intervention might be applied.</p>

which the implementer delivers the program content using the techniques, process, or methods prescribed.” Are the program components delivered as intended? (d) Participant responsiveness, defined as “the extent to which participants are engaged by and involved in the activities and content of the program”; (e) Program differentiation, referring to “whether critical features that distinguish the program from the comparison condition are present or absent during implementation” (O’Donnell, 2008, p. 34).

Researchers in physical education have increasingly recognized the importance of assessing treatment fidelity in intervention research (Escarti, Llopis-Goig, & Wright, 2018; Hastie & Casey, 2014; Kloeppe et al. 2013; Authors, 2021). To advance the literature on treatment fidelity, Authors (2021) adopted the five-element conceptual framework (FOI; Dane & Schneider, 1998; Durlak & Dupre, 2008; Dusenbury et al., 2003; Mihalic, 2004; O’Donnell, 2008), and developed a checklist to assess FOI in physical education. The authors (2021) reviewed and evaluated the quality of fidelity of implementation (FOI) reported in experimental research in physical education pedagogy published in the *Research Quarterly for Exercise and Sport* and *Journal of Teaching in Physical Education*. The findings showed that there were great variations in what was reported in the FOI components in the field of physical education. The FOI information reported by the researchers was limited, incomplete, and vague. Only one study assessed all five components. In addition, FOI is only one dimension of treatment fidelity. Therefore, there is an urgent need to have a conceptual framework to guide intervention research in physical education by including all dimensions of treatment fidelity. A valid and reliable measurement of treatment fidelity is also needed to advance this line of research in the field of physical education.

Currently, no conceptual framework, however, has been proposed to guide treatment fidelity in intervention research in K-12 school physical education. No valid and reliable measurement is available to assess treatment fidelity in K-12 school physical education. In addition, Myers et al. (2021) conducted a review to summarize Measurement in Physical Education and Exercise Science (MPEES)-related activities in 2020 and called for protocol papers for intervention studies to be published in the tutorial/teacher’s toolbox section of MPEES. Therefore, the purposes of this study were to propose a conceptual framework of treatment fidelity to guide intervention research in K-12 school physical education, and then develop and validate a checklist to assess treatment fidelity in K-12 school physical education.

Method

Selecting and Determining a Conceptual Framework of Treatment Fidelity for Intervention Research in K-12 Physical Education

A literature search was conducted on journal article titles using ten key words through EBSCO Discovery Service at the first author’s library. The key words included treatment fidelity, treatment integrity, treatment differentiation, fidelity of intervention, fidelity of treatment, treatment fidelity, implementation fidelity, program integrity, intervention fidelity, and fidelity of implementation. A total number of 5242 articles were identified. Only articles which were published in English, addressed the conceptual framework of treatment fidelity, and included a measurement of treatment fidelity were further reviewed. As a result of the review, we identified two conceptual frameworks which had a corresponding measurement: Conceptual

framework of treatment fidelity in public health (Borrelli, 2011) and conceptual framework of fidelity of implementation in education (O'Donnell, 2008). After comparing those two conceptual frameworks, we made the decision to use both as a basis for the development of treatment fidelity conceptual framework in physical education intervention research, which is described below.

Development of Conceptual Framework of Treatment Fidelity for Intervention Research in K-12 Physical Education

Based on the conceptual framework of treatment fidelity in public health (Borrelli, 2011) and the conceptual framework of fidelity of implementation in education (O'Donnell, 2008), we proposed a conceptual framework of treatment fidelity for intervention research in K-12 school physical education. This conceptual framework consists of five components: Treatment design, training teachers, treatment delivery, receipt of treatment/student responsiveness, and treatment differentiation (See Table 2). The component of treatment enactment in the conceptual framework from Borrelli (2011) was excluded since transfer of learning is an important learning outcome for researchers interested in the impact of curriculum and instruction on student cognitive, affective, and behavioral responses in K-12 physical education. Items from treatment delivery in the conceptualization of Borrelli (2011) were similar to those proposed by O'Donnell (2008). Therefore, we used all the items and indicators from O'Donnell (2008) for treatment delivery since they are more applicable to research in educational settings. Receipt of treatment (Borrelli, 2011) was replaced by participant responsiveness (O'Donnell, 2008) since they are similar concepts.

Development of Treatment Fidelity Checklist for Intervention Research in K-12 Physical Education

Two senior researchers who have expertise in statistics carefully read the Bellg et al. (2004), Borrelli et al. (2005) and Borrelli et al. (2011) articles to gain understanding of the conceptual framework of treatment fidelity and Treatment Fidelity Checklist. These three articles were selected because they were original articles which provided detailed descriptions of the 5-component conceptual framework of treatment fidelity, explanations of Treatment Fidelity Checklist, and practices and strategies for using Treatment Fidelity Checklist.

Guided by our proposed conceptual framework, the two senior researchers developed the Treatment Fidelity Checklist through the modification of the original 29-item checklist published by Borrelli (2011). By going through each of 29 items in the Treatment Fidelity Checklist (Borrelli, 2011), we operationalized the definition of all the items and made modifications to adapt to the physical education teaching and learning context in schools. Items were deleted if they are not applicable or common practices which are typically not reported in published articles. Such modification resulted in the deletion of several items. The item "method to ensure that dose is equivalent for participants within conditions" was deleted since participants will receive differentiated, individualized instruction in teaching and learning. As a result, the dose may vary. It is a common practice to have a plan to address possible setbacks in implementation in field-based research on teaching and learning. The researchers do not report this information in their published articles. Therefore, this item was deleted. The item "At the hiring stage, assessment of whether or not there is a good fit between the provider and the intervention (e.g.,

Table 2
Fifteen-Item Treatment Fidelity Checklist

Components/Items	Indicators	Ratings		
		Present	Not-present	Non-Applicable
Treatment Design				
1. Provide information about treatment exposure in the intervention condition	1. Length of contact	How many minutes per physical education lesson?		
	2. Number of contacts	How many lessons per week?		
	3. Content of treatment: What was manipulated?	Intervention program content		
	4. Duration of contact over time	How many weeks or semesters will the intervention program last?		
2. Provide information about treatment exposure in the comparison condition or control condition	1. Length of contact	How many minutes per physical education lesson?		
	2. Number of contacts:	How many lessons per week?		
	3. Content of treatment: What was manipulated or what was taught in the control group?	Lesson plans/program content for the comparison or control group.		
	4. Duration of contact over time	How many weeks or semesters will the intervention program last?		
	5. Method to ensure that exposure is equivalent between conditions	1. Classes are sampled from the same schools or school districts to ensure the equivalent treatment.		

2. Make-ups for any missing lesson.
 3. Record keeping of intervention exposures
 4. Modify lesson plans to make up any deviations from the planned intervention exposures
- At least one indicator reported to be coded as “present.”
- Providers’ background and training history, teaching experiences, specialist teacher versus non specialist, education, certificates/honors, etc: If any information is provided, it will be coded as “present”. particularly, Credentials of teachers that researchers need for their studies.
- The intervention components are aligned with the constructs of theoretical model.
1. Use a well-established valid and reliable measure and
3. Specification of teachers’ credentials relevant to the intervention that are needed. Characteristics being sought in the teachers are articulated a priori. Characteristics that should be avoided in the teachers are articulated a priori.
4. Theoretical or conceptual model or empirical work upon which the intervention is based is clearly articulated
- The active ingredients are specified and incorporated into the intervention
- Measures reflect the hypothesized theoretical constructs/mechanisms of action

	provide a citation for it. At least the reliability of the current study data is reported.
5. Potential confounders that limit the ability to make conclusions at the end of the trial are identified	2. Use a self-created measure but the reliability and validity of it are tested and reported in the current study. Potential confounders are identified and/or controlled in the design and/or statistical analysis. E.g., teachers' characteristics, sampling bias, covariates, etc.
6. If more than one condition (including control/comparison condition) is described, all described equally well	All conditions are described equally well.
Training Teachers	Manual or written document of training procedures
1. Description of how teachers will be trained for the intervention.	Explicitly indicating that training manual are used and followed across multi-sites/teachers.
2. Standardization of teacher training, especially if multiple waves of training are needed for multiple groups of teachers.	Use knowledge
3. Assessment of teacher's skill acquisition	

International Journal of Physical Activity

		test/teaching demonstrations or other measures to assess how well the teachers are trained.
4. Assessment and monitoring of teachers' skill maintenance over time		Feedback provided during training/retraining/re-testing and during the implementation of the study
Treatment Delivery		
1. Program Adherence	Program intervention components being delivered as designed	Use logs, diaries, observation instrument, checklist, videotaping, audio, computer software to check that the content of the intervention is delivered as specified
2. Duration	Method to ensure that the dose of the intervention is delivered as specified	The number, length, or frequency of sessions implemented: Reporting attendance data from each participant
3. Delivery Quality	The manner in which the implementer delivers the program using the techniques, process, or method prescribed	Program delivery quality data (field notes, measurements, checklist, observational instruments) are reported
Receipt of Treatment/Student Responsiveness		
	The degree to which students understand	Use quantitative data such

International Journal of Physical Activity

the intervention and the extent to which participants are engaged and involved in the activities and content of the program

as surveys and tests or qualitative data such as interview or field notes to measure how well students understand the intervention and how much students are involved and engaged in the program beside the dependent variables of interest.

Treatment Differentiation

Critical features of intervention and comparison/control programs

Whether critical features that distinguish the program from the comparison condition are present or absent during implementation: two conditions at least. One condition: N/A

ensure that providers find the intervention acceptable, credible, and potentially efficacious)” was deleted since the fit between the provider (typically physical education teachers) and the intervention was addressed when conducting recruitment and signing consent form. However, researchers rarely reported this information in their published articles. The items “assessment of nonspecific treatment effects” and “there is a plan for the assessment of whether or not prescribed components were delivered. (e.g., components that are unnecessary or unhelpful)” were deleted since they are not applicable to research in the field of teaching and learning in physical education. Many items from receipt of treatment by Borrelli (2011) were also deleted since no matter what interventions are, teachers do use pedagogies to improve student understanding of interventions and their skill performance.

The modified Treatment Fidelity Checklist contains the list of criteria (18 items) by which articles were evaluated. The 18 items are divided into the five treatment fidelity categories (Treatment Design, Training Teachers, Treatment Delivery, Receipt of Treatment/Student Responsiveness, and Treatment Differentiation). Examples were provided for each of 18 items to help the coders determine whether the article met the criteria for that particular item. Coders rated the presence or absence of a number of attributes within each of the five categories, which are important for the preservation of treatment fidelity. If the treatment fidelity information provided in the article met the criteria, it was coded “present” with a number “1” for that particular item. If the treatment fidelity information provided in the article did not meet the criteria, it was coded “absent but should be present” with a number “0” for that particular item. If the particular treatment fidelity information was not applicable to the article, it was coded as “non-applicable” with a number “2” for that particular item. For example, for the item “If more than one intervention is described, all described equally well”, articles with only one intervention were coded as “non-applicable” with a number “2” for this item. Both coders used the modified Treatment Fidelity Checklist to code one article together. Discrepancies were discussed and a 100% consensus was reached.

Validation of Treatment Fidelity Checklist for Intervention Research in K-12 Physical Education

The validation of treatment fidelity checklist was conducted in two phases. In the first phase, four experts who published research work on treatment fidelity and are also proponents of treatment fidelity were recruited to examine the content validity of this checklist. In the second phase, two researchers used the checklist to code at least 33% of published intervention research (14 articles), which reported fidelity check data to establish intercoder reliability.

The four experts were asked to complete a survey. The survey consisted of two parts. The first part included detailed information related to two conceptual frameworks, how we integrated both frameworks, and how we developed the checklist. The second part included seven questions for experts to complete. Those questions were: (a) Overall, how do you perceive this proposed conceptual framework of treatment fidelity for intervention research in K-12 physical education? (b) Please rank the importance of the five categories and their corresponding indicators on 1-5 point Likert scale: 5 = most important, 4 = important, 3 = neutral, 2 = not important, 1 = not important at all; (c) Please write down any categories and their corresponding indicators that you would like us to add below; (d) Please write down any categories and their corresponding indicators that you would like us to delete below; (e) Please write down any modifications that

you recommend us to make for those categories and their corresponding indicators; (f) How do you perceive our use of ratings of “present”, “Not-present”, or “non-applicable” to code the presence or absence of a number of attributes within each of the five categories? and (g) Do you have any additional comments?

In addition to requesting clarifications on some items, the experts also made many comments. Those comments are presented in Table 3. We made modifications and clarifications to the checklist based on the experts’ comments. Since statistical design is not part of treatment fidelity, it was not included in the treatment design. We also changed study design to treatment design to avoid any confusion. Since there are a lot of variations in teacher training with some being a couple hours, we did not align the teacher training dimension with principles and practices of effective professional development. For the rest of experts’ comments, we address them in the sections of discussions and limitations. The final fifteen-item Treatment Fidelity Checklist is provided in Table 2. The fifteen items are all important for the preservation of treatment fidelity. Each of them is coded “present”, “absent but should be present”, or “non-applicable” with corresponding numerical values: 1, 0, and 2.

Table 3
Experts’ Comments

-
- 1) It would be important for researchers to be able to indicate when aspects of the framework may not apply to a particular study. For example, not all intervention studies in our field have a comparison condition.
 - 2) It is not always possible to make up for missing dosages in field research (but this should be reported). Perhaps use a different term than dosages, such as, “lessons” or “sessions”?
 - 3) Sometimes it is better to not sample from the same school due to contamination.
 - 4) The guiding theory or conceptual theory does not always directly translate to intervention components.
 - 5) Experts are not always surveyed on protocol.
 - 6) The potential confounders seem like there may be data analysis issues.
 - 7) Description of training is important – I’m not sure it needs to be a “manual” perhaps just “written”.
 - 8) I’m not sure that researchers should comment to teachers on their skill maintenance over time.
 - 9) The training plan issues would depend on the scale of the study.
 - 10) For receipt of treatment, there may be some equitable qualitative measures that could also be used.
 - 11) As long as the researcher can indicate non-applicable (without penalty), it should be ok to keep categories.
 - 12) I like the flexibility of using ‘at least one indicator being reported to be coded as “present”’ that was used in the first section of the instrument. Can it be used throughout the instrument?
 - 13) I think the tool could be really useful for researchers in planning and reporting studies (e.g., including limitations); however, research conducted in our field would rarely meet all or even most of these indicators.

- 14) A design issue that goes beyond treatment fidelity or fidelity of implementation that might be applicable under “treatment design” that is not indicated in the instrument (that members of our field often struggle with – myself included) is the unit of analysis (e.g., class or school vs. individual participants).
 - 15) This seems a little arbitrary. Also, ethnicity has nothing to do with credentials. Indicators may need to be relevant to credentials.
 - 16) I wonder whether there’s a need for some guidance / information about an overall fidelity score / rating?
 - 17) I think it is critical that any teacher training or professional development attempt is (explicitly) aligned with principles of effective professional development.
 - 18) It seems to me that if a fidelity checklist must address the extent to which elements are completed and the extent to which the elements have some magnitude measurement in the actual study. For example, if a study says that they provided feedback some judgement must be made from a fidelity perspective.
 - 19) Is each element equal or are some more than others?
 - 20) How do you judge a study that doesn’t have all the information in it because of page limitations in the journal this could cause over and underestimation effects depending on how you interpret what you see. For example, whether a plan for how contamination between conditions will be prevented- I have used such efforts in every study I have done and have never reported it.
 - 21) I think that if you use this checklist, you will find most if not all studies will be viewed poorly because the number of elements seems a bit over the top.
-

Intercoder Reliability. In the second phase, two researchers used the checklist to code at least 33% of published intervention research which reported fidelity check data to establish inter-coder reliability. Both researchers worked as a rank of Professor at universities in the United States. They are senior researchers who have a PhD degree in sports pedagogy, have a significant record of quality publications in their field, and have great expertise in statistics.

This study focused on experimental research in K-12 school physical education, which reported treatment fidelity check and was published in the *Journal of Teaching in Physical Education* and in the pedagogy section and Research Notes on Pedagogy topics of *Research Quarterly for Exercise and Sport* from January 1998 to December 2020. Nine articles were randomly selected from the database reported by Authors (2021), which has 35 articles published from January 1998 to December 2018. Eight additional articles published from January 2019 to December 2020 that met the inclusive and exclusive criteria were identified by the first author. Five out of eight articles were randomly selected. A total number of randomly selected sixteen studies were included for this study for intercoder reliability check.

To familiarize with the coding template, two articles were randomly selected for coding. Both coders independently coded the articles using the modified Treatment Fidelity Checklist. Discrepancies were discussed and a 100% consensus was reached. Some operational definitions were further clarified. Then both coders independently coded the rest of fourteen articles. The intercoder reliability was established by using the percent agreement between the two coders (Lombard, Snyder-Duch, & Bracken, 2002).

Results

All four experts agreed that the development of conceptual framework of treatment fidelity for intervention research in K-12 physical education was really a good idea from a scientific perspective and could be a very useful guidance document. The framework and measurement of treatment fidelity were validated by experts. The overall IOA was 88% between the coders.

Discussions and Conclusion

In the present study, we proposed a conceptual framework of treatment fidelity to guide intervention research in K-12 school physical education based on the conceptual framework of treatment fidelity in public health (Borrelli, 2011) and conceptual framework of fidelity of implementation in education (O'Donnell, 2008), and then developed a checklist (i.e., Treatment Fidelity Checklist) to assess treatment fidelity in K-12 school physical education. Our proposed conceptual framework is more comprehensive and applicable to intervention research in school physical education settings by integrating both frameworks from public health and general education. All four experts recognized and valued the importance of this proposed conceptual framework. The importance of developing this conceptual framework is also a timely and urgent response to the increasing call from researchers in physical education to measure the fidelity of implementation of intervention. The validation data demonstrated that the Treatment Fidelity Checklist is a valid and reliable measure to assess treatment fidelity in K-12 school physical education. The conceptual framework and checklist of treatment fidelity provide a very useful document to guide future intervention research in the field of physical education and what to be reported in publications.

However, based on the experts' comments, several limitations shall be taken into considerations when using this checklist. First, there are a lot of elements on this checklist. Those elements are critical to the treatment fidelity of an intervention study. However, due to page limits, researchers may not report all of them in their publication. In the future, a flexible approach shall be used when researchers use this document to guide their report of research. Secondly, for some elements on this checklist, the extent to which elements are completed was addressed. However, for other elements, the magnitude measurement was not applicable. Therefore, an overall fidelity score/rating may not be feasible and practical. Third, all elements are treated as equal. There is no research evidence available to support an assignment of weight to different items according to their importance. It may not be possible to assign a differentiated weight to all the elements.

The present study took an initial step to propose a conceptual framework and then developed a valid treatment fidelity checklist for intervention research in the field of K-12 physical education. A further examination and validation of the Treatment fidelity checklist is warranted by addressing those limitations.

References

- Bellg, A., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., et al. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH behavior change consortium. *Health Psychology, 23*, 443–451.
- Borrelli, B., Sepinwall, D., Bellg, A. J., Breger, R., DeFrancesco, C., Sharp, D. L., Ernst, D., Czajkowski, S., Levesque, C., & Ogedegbe, G. (2005). A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *Journal of Consulting and Clinical Psychology, 73*(5), 852-860.
- Borrelli, B. (2011). The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. *Journal of Public Health Dentistry, 71*, S52-S63.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*, 23–45.
- Durlak, J. A., & Dupre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on programs outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327–350
- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school setting. *Health Education Research, 18*(2), 237-256.
- Escarti, A., Llopis-Goig, R., & Wright, P. M. (2018). Assessing the implementation fidelity of a school-based teaching personal and social responsibility program in physical education and other subject areas. *Journal of Teaching in Physical Education, 37*, 12-23.
- Hastie, P. A., & Casey, A. (2014). Fidelity in models-based practice research in sport pedagogy: A guide for future investigations. *Journal of Teaching in Physical Education, 33*, 422-431.
- Kloeppel, T., Hodges Kulinna, P., Stylianou, M., & van der Mars, H. (2013). Teacher fidelity to one physical education curricular model. *Journal of Teaching in Physical Education, 32*, 186–204
- Li, W., Ma, L., Xiang, P., & Tang, Y. (2021). A review of fidelity of implementation in intervention research published in the *JTPE* and *RQES*. *Journal of Teaching in Physical Education, 40*, 662-666.

International Journal of Physical Activity

- Li, W., Xie, X., Li, Y., Chen, R., & Shen, H. (2021). Twenty years of intervention research in school physical education in the *JTPE* and *RQES*, *Research Quarterly for Exercise and Sport*, *92*, 34-42.
- Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communications Research*, *28*, 587–604.
- Mihalic, S. (2004). The importance of implementation fidelity. *Emotional & Behavioral Disorders in Youth*, *4*, 83–86 and 99–105.
- Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, *24*, 315-340.
- Myers, N.D., Bateman, A.G., Lee, S., & Silverman, S. (2021). Measurement in Physical Education and Exercise Science (MPEES): A summary of MPEES-related activities in 2020. *Measurement in Physical Education and Exercise Science*, *25*, 191-201.
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, *78*, 33-84.