

QUANTITATIVE RESEARCH TOOL EXAM

Critical Analysis:

Wirkala, C. & Kuhn D. (2011). Problem-based learning in K-12 education: Is it effective and how does it achieve its effects? *American Educational Research Journal* 48(5): 1157 - 1186

Presented by:

Lorraine Emerick

University of New York at Albany

Examination Committee:

Dr. Peter Shea, Chair

Dr. Joseph Bowman, Examiner

Critical Analysis: “Problem-Based Learning In K-12 Education: Is It Effective and How Does it Achieve its Effects?”

Part 1 Description of the Overall Study

It is vitally important that the curricular offerings being delivered today achieve their goals of providing young people with high-quality evidence-based educational services that will prepare them for life in the 21st century. In their recent study, “Problem-Based Learning in K–12 Education: Is It Effective and How Does it Achieve its Effects?” (2011), Wirkala, a PhD in developmental psychology, and Kuhn, a professor of psychology and education, report that despite the widespread popularity of problem-based learning (PBL) among educators and growing perceptions of its status as “best practices,” there remains insufficient rigorous experimental evidence concerning its effectiveness, particularly among K–12 pupils and students. As defined by Wirkala and Kuhn, “problem-based learning (PBL) is a teaching and learning method in which students engage a problem without preparatory study and with knowledge insufficient to solve the problem, requiring that they extend existing knowledge and understanding and apply this enhanced understanding to generating a solution” (2011, p. 1157).

The researchers used a highly controlled post-test only experimental study design of different PBL applications in a middle school population. According to Wirkala and Kuhn, this high degree of experimental control within a classroom setting is considered a major strength of the study whereby “students learned in their familiar, real-world school setting, enhancing external validity, while the experimental design maximized internal validity, with instructor, curricular objectives, content, and schedule of instruction equated across conditions” (2011, p. 1181). The study was guided by the overarching research question, “do the benefits of PBL justify its demands?,” which was answered by breaking the multifaceted PBL down into its

constituent elements and evaluating the respective effectiveness of each using a theoretical framework based on earlier work by Capon and Kuhn (2004) and Pease and Kuhn (2011). These studies were also used to evaluate the effectiveness of PBL in natural instructional settings under tight experimental control. In contrast to the other two studies, Wirkala and Kuhn state that, “we compare not only PBL and lecture/discussion (LD) instructional conditions but also two forms of PBL instruction—team and individual—in order to examine whether the effectiveness of PBL is reduced when its social component is subtracted, and hence whether social collaboration is an essential component of the PBL method” (p. 1159). The research design and analysis used to achieve this comparison was a crossed within-subjects design that manipulated two independent variables: (a) instructional format (PBL vs. LD); and, (b) grouping condition (PBL-team vs. PBL-individual). Based on the study’s findings, Wirkala and Kuhn conclude that compared to the traditional lecture method, PBL was in fact superior in helping students achieve mastery, but that the social component does not contribute to its overall effectiveness.

Part II. Critical Evaluation

Although Wirkala and Kuhn did not elucidate a specific research question or questions per se, they did explain the overall purpose and general question that guided their research early on in their content. For instance, although they state that, “in order to address the research questions of specific interest in this study, analyses for most performance variables consisted of two planned comparisons—one between the LD group and the PBL groups and one between the two PBL groups” (p. 1172), they do not ever articulate the specific research questions that guided their study. These researchers, though, do report that, “given its growing use, and the potential for much more widespread use, in K–12 education, the question of whether its benefits justify its demands is thus one of great practical significance” (p. 1159). Utilizing a positivist

philosophical stance, Wirkala and Kuhn support their choice of research design, population, setting, and data analysis techniques, by articulating, “we see the present study as contributing to an essential research base necessary to answer this question” (p. 1159). Until now, the majority of study participants have been adults in university settings. There are almost always ethical issues raised when using human subjects, especially children, such as the middle school participants in this study. While this complicates the ethical climate, it does not in and of itself impact upon the integrity of the report, however, Wirkala and Kuhn do not divulge the IRB rigor inherent in research involving human subjects.

Researchers present a fairly comprehensive review of the recent relevant peer-reviewed and scholarly literature and provide a synthesis of the main themes as they apply to their research. With respect to their choice of design, selection of sites and/or participants, and data collection and analysis techniques, Wirkala and Kuhn advise that the research design covered two basic topics as follows:

1. Topic 1 was groupthink, the faulty decision making that can occur in groups with low cognitive diversity and other characteristics.
2. Topic 2 was learning and memory, particularly how certain study factors affect memory for learned material (p. 1160).

All of the subjects were instructed in both topics using the PBL or the lecture/discussion (LD) approach. The instruction for each topic took place during three 40-minute class sessions, over the course of 1½ weeks (two hour total). Entire intervention, including both topics, consisted of 6 class sessions and a total of 4 hours. Groupthink occurred at the end of the sixth-grade year; instruction on the second (memory) at the beginning of the seventh-grade year (separated by summer break). Long-term learning was assessed 9 weeks after instruction ended -

via cued assessment of comprehension and an uncued assessment of application (to a new context). It should be noted that the researchers were not interested in short-term gains but endured learning (Wirkala and Kuhn, 2011). According to Wirkala and Kuhn, “across topics, 24 to 31 students were included in the between-subjects analyses, and 19 to 22 were included in the within-subjects analyses” (p. p. 1173).

Wirkala and Kuhn established a baseline in an effort to understand the targeted concepts and the equivalency of the two topics (groupthink and learning/memory) by utilizing an additional group of 94 students from one grade below at the same school. The 94 students were administered the cued comprehension assessments, however, they did not participate in study (2011, p 1160). Wirkala and Kuhn (2011) also stated, but did not elaborate further,

"that the content was developmentally and age appropriate for the population. However, it was entirely new to these students and different from the typical content in the course in which the intervention took place (social studies), thus minimizing previous knowledge as a variable factor across participants" (p. 1160).

This aspect of their study, they suggest, helps mitigate prior knowledge as an interfering variable for all subjects. While these measures helped to ensure the reliability and trustworthiness of the study’s findings, the authors do not present a discussion concerning alternative methods that could have provided additional insights nor do they attempt to explain or interpret their findings differently based on the clear-cut data that resulted from the statistical analyses.

According to Wirkala and Kuhn, the statistical analyses and the rationale in support of their use in this study are set forth in Table 1 below:

Table 1
Statistical Analyses Used by Wirkala and Kuhn (2011)

Type	Description*	Authors' Rationale**
t test	The t test is a parametric test (<i>mean</i> that is calculated from data and describes a population) and is used to determine if two means are significantly different at a selected probability level.	This test was used to evaluate the difference in mean number of concepts defined/applied.
chi-square	This tests the hypothesis that variables are independent, without indicating their strength or direction of the relationship. Chi-square is a non-parametric test (no dependency on any parameters) often used when data are in the form of frequency counts, percentages...that can be converted to actual numbers.	The chi-square statistic was used where ordinal scales were involved to assess depth of explanation achieved.
Wilcoxon signed-rank test	A nonparametric procedure used with two related variables to test the hypothesis that the two variables have the same distribution. This test takes into account information about the magnitude of differences within pairs and gives more weight to pairs that show large differences than to pairs that show small differences.	The Wilcoxon signed-rank test was used to assess individual patterns over the two topics (groupthink/ learning & memory).

Sources: * Kaufhold (2007); ** Wirkala and Kuhn (2011)

Based on the test description and supporting rationale provided by the researchers, it appears that the statistical analyses were appropriate for the purposes for which they were intended, but the authors do not include any specific justification for the use of these statistical tests beyond their intended purpose. The authors' description and explanation of their research design, though, required careful reading to fully comprehend, a process that laypersons might be reluctant to pursue.

The fundamental ethical issue raised by this study concerned the author's assumption that problem-based learning represents an across-the-board, end-all "best practices" pedagogical solution rather than a valuable alternative that must be used in those situations where it works best. For many educators, it is axiomatic that if students are not learning the way they are taught, they must teach the way students learn. The assumption concerning the best practice status of

practice-based learning by Wirkala and Kuhn does little to illuminate the role that PBL can play the digital age, or in educational settings that are becoming increasingly multicultural.

Wirkala and Kuhn seek to overcome the weaknesses that are inherent in the previous studies concerning the effectiveness of PBL, including “(a) nonrandom assignment of students to PBL and traditional instruction, (b) variations in time and exposure to treatment across frequently lengthy interventions, and (c) varying instructors and conditions, undertaking to study the effectiveness of PBL in a natural instructional but highly controlled experimental educational setting” (2011, p.1158). To this end, the authors emphasize that, “we also follow their approach in acknowledging the varying practices that have been characterized as falling under the heading and undertaking to instantiate PBL in its ‘best practice’ form, namely, as its advocates claim it to be most effective” (p. 1159). While the researchers claim to not have the information about the effectiveness of PBL versus traditional LD delivery modalities, the aforementioned suggests innate bias against LD. While the site/participant selection is and does relate well (subjects appear diverse and valid), the data collection and analysis methods, conclusions and interpretations do not necessarily relate to the research questions and theoretical framework due to its innate bias against LD and the lack of an outside control group. It should also be noted that Wirkala and Kuhn (2011) claim that the superior performance of their approach might have been simply due to some students in the group “possessing the needed expertise” (p. 1183). Unfortunately, the authors did not test for this possibility at the outset - in other words, some students may be coming in better equipped to succeed at PBL.

Other limitations of the study include the idea that "groupthink" might actually constitute consensus building – specifically with regard to competitive types of learning. In addition to male/female physiological differences not considered (gender equality) nor were individual

learning styles. Other interpretations of the results are possible, as Wirkala and Kuhn (2011) admit when they point out:

"Still, we cannot rule out the possibility that the instructor consciously or unconsciously delivered a superior product in one case due to subtle differences that were extraneous to the definitions of each practice. In the case of PBL practice, this potential influence is diluted by the presence of multiple coaches and the indirect role they play in instruction" (p.1181).

The instructor who delivers a superior instruction due to extraneous "subtle differences" may have delivered an instructionally better product due to the fact they that they were simply better prepared or more knowledgeable. The multiple coaches in PBL will then dilute this advantage. The unmentioned result may be that the LD product is better by default.

Another critical issue that Wirkala and Kuhn do identify as an area in their study that needs more research is the number of coaches used in PBL. This may be a critical area of change that needs to be modified in order to make the modality work more effectively by itself or in coordination with LD. The study's PBL groups had two extra coaches on staff who supplemented the primary teacher. Unfortunately, employing extra coaches is not always realistic in a school setting due to budget constraints.

As to the contributions these findings made to theory development and future practice or research, as well as the conclusions and interpretations derived from their findings, Wirkala and Kuhn (2011) note "although the present work focuses on outcomes rather than process, we believe our findings indicate that the more laborious process observations and analyses warrant the investment" (p. 1184). The implications of the study and the recommendations for further

research are congruent with the findings that emerged, but beyond the social component issue, few new insights were developed during this research project.

Nevertheless, for many educators (as well as business managers and governmental policymakers), it has also become axiomatic that in order to improve something, relevant metrics and benchmarks must first be established and steps taken to evaluate the effectiveness of an initiative and the study by Wirkala and Kuhn showed that this is certainly the case with problem-based learning scenarios. After all, schools are not so many mindless factories churning out little widgets to join society when they are completed, but are rather the focus of an enormous amount of taxpayer resources that demand efficient and effective applications. Because young minds are involved, there is little room for experimentation or false starts, and it is important to use the pedagogical approaches that have demonstrated effectiveness and to demonstrate how and why these methods are effective so they can be replicated elsewhere.

Most experienced educators would likely agree from the outset that young people learn better when they are actively engaged in the learning process and have opportunities to use hands-on problem-solving approaches in their classrooms. Indeed, the authors concede the seeming redundancy of their study by noting, “the evidence presented here conceivably will be seen by some as showing little more than what is taken as a given on the part of a good number of practitioners: Students show better long-term retention and ability to apply new material if the instructional method is one that actively engages them and enables them to put new ideas to use” (p. 1180). Clearly, at this point, the researchers are preaching to the choir, but they are quick to add that these findings represent more than just another academic exercise that should be filed away with the other studies concerning how problem-based learning helps students achieve improved academic outcomes. In this regard, Wirkala and Kuhn (2011) stress, “we see the

present research as more than a rigorous demonstration of the obvious. Rather, we see it as a starting point, not a final seal of approval, on a path toward the important goal of solidly evidence-based instructional practice” (p. 1180).

At this point, the researchers jump on a very slippery slope by suggesting that yet further research into the constituent components of PBL and how it achieves its effects is warranted by the research to date. For instance, Wirkala and Kuhn conclude that the question as to whether the benefits of PBL can be used to achieve improved academic outcomes by affecting student disposition toward learning demands even more research, an issue they maintain is especially relevant for very young learners. According to Wirkala and Kuhn (2011):

“Especially for the K–12 population, this is a critical question. It is one, however, that can only be answered with continuing rigorous investigations of both processes and products of PBL, including investigation of students’ own assessments of their learning experience” (p. 1184).

The main findings of this study were that both versions of PBL were far superior in long-term learning lecture/discussion sessions and that there was no difference between PBL-individual and PBL-team situations. This indicates that “social interactions are not what makes PBL so effective” (Wirkal & Kuhn, 2011, p. 1157). Researchers stated their conclusion was general ... collaborative educational methods yield no benefit and "social interaction by itself is not a "magic bullet" that benefits students" (Wirkal & Kuhn, 2011, p. 1183). They countered it by stating that until "microgenetic observations of collaborative learning are carried out, we are limited in what we can conclude about its nature." Considering the small scale of the study (small number of groups involved), is it justified to state that the social component is not

significant? From the limited information measured on post-tests, the social component does not appear to be analyzed or observed in depth.

Wirklan and Kuhn (2011) also addressed the strengths of the short period, small scale of the study with respect to internal validity due to shorter interventions than the typical 1 year time span found in PBL literature as well as highly defined learning objectives. However, the limitations of the design of the present study cannot readily be replicated with instruction of semester-long duration without introducing other variables. Wirkala and Kuhn (2011) acknowledge the best strategy would be to "very gradually increase length of instruction (as well as to vary other factors such as subject matter and grade level) to ascertain how broadly the present findings extend" (p. 1181).

We live in an increasingly collaborative and connected world and educators continue to incorporate computer-based applications in the classroom, as well as various one-to-one laptop initiatives, that are changing the way educational services are delivered. Problem-based learning can certainly be applied to these new learning resources (i.e. simulated virtual environments conducive to situated learning involving problem based scenarios), but because every classroom is unique, a better use of scarce resources might be to investigate how these current trends will affect PBL. Additionally, more research is needed with regard to training teachers in their jobs as PBL coaches. Like any modality of instruction, PBL outcomes depend most heavily on the skills of those who implement them (Wirkala and Kuhn, 2011, 1183). The best tool kit is one that consists of not just one method, but inclusive of the best of PBL and traditional methods combined and used in the classroom together, adapted according to local student needs.

References

- Capon, N., & Kuhn, D. (2004). What's so good about problem-based learning? *Cognition and Instruction*, 22(1), 61–79.
- Kaufhold, John A. (2007). *Basic statistics for educational research*. Lincoln: iUniverse.
- Pease, M., & Kuhn, D. (2011). Experimental analysis of the effective components of problem-based learning. *Science Education*, 95, 57–86.
- Wirkala, C. & Kuhn, D. (2011, October). Problem-based learning in K–12 education: Is it effective and how does it achieve its effects? *American Educational Research Journal*, 48(5), 1157–1186.