

The Impact of Teacher Preferences in Learning by Evaluating

Scott R. Bartholomew, Brigham Young University

scottbartholomew@byu.edu

Emerson Barnum, Brigham Young University

emersongrace2002@gmail.com

Andrew Jackson, University of Georgia

andrewjackson@uga.edu

Nathan Mentzer, Purdue University

nmentzer@purdue.edu

Jessica Allen, Brigham Young University

jessica.yauney@gmail.com

ABSTRACT

Peer review and the evaluation of samples are common tools used in education and our research has shown positive impacts on student learning through the intentional evaluation of samples as a priming exercise—an experience we have termed “Learning by Evaluating (LbE).” While previous work in design and technology (D&T) classrooms with LbE has demonstrated positive student learning gains, we have not yet investigated the impact of the classroom teacher on the LbE experience for students. Therefore, our research sought to analyze the impact of a classroom teacher on student experience within a LbE experience situated in a D&T classroom. To better understand the impact the teacher has during LbE, multiple D&T classrooms engaged in LbE sessions were enrolled in this study. Each D&T classroom followed a similar protocol which had students engaged in an open-ended design problem who used LbE to evaluate samples of related work as part of their designing process. Specifically, we collected data from student LbE decisions to explore if students in different D&T classes valued different elements of the samples (e.g., did students in one class focus on aesthetics while students in another emphasized brevity in explanations). An online software platform (RMCompare) was used to engage students in LbE and collect both the quantitative data associated with the ranked preferences of the students and the qualitative data from their justifications for their selections. All students were enrolled in the same district-level course and presented with the same samples. Key findings—both similarities and differences—between classes will be shared in alignment with implications for design and technology classrooms.

Key Words: Adaptive Comparative judgment, Learning by Evaluating, Teacher Impact

1. INTRODUCTION

Research and investigation around the potential to use assessment and evaluation as a learning activity has led to a growing body of evidence (Bartholomew, Mentzer, Jones, Sherman, & Baniya, 2020) around the idea that students can learn by evaluating (LbE). For example, K-12 (Bartholomew & Yoshikawa-Ruesch, 2018) and university students (Bartholomew & Yoshikawa-Ruesch, 2018) have engaged in LbE—typically facilitated by an online software and process called adaptive comparative judgment—with positive learning gains and improvements to the overall design process. In LbE the act of evaluation is situated near the beginning of a learning cycle (e.g., the design process) and used as a primer for students’ later experience and work. Whereas traditional classroom approaches typically place evaluation at the conclusion of an assignment (with teachers as the evaluators), LbE places it at the forefront and engages students in the process of the evaluation.

Lacking in the present research around LbE is an exploration of the impact, if any, different teachers have on student experience when doing LbE (Bartholomew & Jones, 2020). For example, is it possible that differences in teacher style, classroom management practices, etc. may impact the way students engage in LbE? Further, how do these differences shape the student learning and experience? Given the rise in the use of LbE, and the positive findings for student learning, it is important to explore these ideas and the ways teachers may, or may not, influence the LbE experience. Our stated research question was:

How are student LbE experiences similar, or different, as a result of classroom-teacher differences?

2. LITERATURE REVIEW

Although not required, LbE to date has largely been facilitated as part of a larger adaptive comparative judgment (ACJ) experience for students (Bartholomew & Jones, 2020). Therefore, a correct understanding of the current research into LbE will not be complete without a basic review of ACJ, its theoretical basis, and the way LbE is facilitated through ACJ.

2.1. *Adaptive Comparative Judgment*

ACJ is an approach to assessing the quality of items through comparison rather than subjective judgment, value or point-based allocation, or traditional rubrics (Kimbell, 2021). In an ACJ setting, an individual—or group of individuals—views pairs of items (e.g., design journals, product pitches, essays, interviews, etc.) and selects which item—of those displayed—they believe better satisfies a predetermined criterion. Following each selection, a new pair of items is displayed (which may or may not include one of the items previously displayed), and another comparative decision is made. These comparative decisions (a.k.a. judgments) are made until a rank order of all items is created – often with very high levels of reliability (Bartholomew & Jones, 2020). In addition to the rank order, parameter values can be generated as part of the process—parameter values differ from the rank in that they show both a magnitude and direction (e.g., ranks are all equidistant but parameter values denote both the direction of change and the

magnitude/size of that change); these values are influenced by Rasch-modelling misfit statistics (see Pollitt, 2004 for a full discussion of parameter values). Finally, judge rationale for decisions are often collected as part of the comparative process through an automated software or other means.

Originally, ACJ was put forth as an alternative to traditional approaches to assessment which could be time-consuming and highly unreliable (Bartholomew, Strimel, Zhang, & Homan, 2018). However, research has shown that when students use ACJ as a learning primer (rather than a teacher using ACJ for assessment of student work), student learning is positively impacted (Mentzer, Lee, & Bartholomew, 2021). Further, research has shown high reliability levels even between student and teacher judgments (Baniya et al. 2019, Bramley 2015), an “easier” assessment process (Kimbell, 2021), and applicability in facilitating feedback from multiple assessors (Bartholomew and Yoshikawa 2018; Kimbell 2012b).

2.2. Learning by Evaluating

In 2020 Bartholomew, Mentzer, Jones, & Sherman coined the acronym “LbE” to describe the process of students viewing and evaluating examples of work using Adaptive Comparative Judgment (ACJ) prior to engaging in similar assignments themselves. Their work built on previous research into student ACJ for learning (Baniya et al. 2019; Bartholomew and Strimel 2019; Bartholomew et al. 2018a; Bartholomew et al. 2018b; & Seery and Canty 2017). Student comments about LbE have highlighted its ability to help them gain confidence (Canty 2012) and improve their own work (Bartholomew et al. 2019) as benefits of using LbE in the classroom. LbE has been used in various fields including Design, English, Engineering, and Business (see Bartholomew & Jones, 2020).

2.3. Teacher impact on student experience

Traditionally, LbE (and ACJ) research has not explored the similarities and/or differences in student experience based on class and/or teacher; rather, these studies have focused largely on treatment-level conditions (control or treatment group)—students who engaged in LbE and those that did not. However, academic studies have consistently shown that the teacher has the single biggest impact on student learning over any other variable (Hattie, 2015; Rockoff, 2004). This has held true despite differences in lessons, unit, school, location, and a variety of other factors (Hattie, 2017). As Fountas & Pinnell (2023, n.d.) note:

Though a teacher may lean on a lesson to deliver instruction, the teacher is always the most critical factor in determining what a child can achieve in the classroom. No lesson plan or program will singlehandedly identify and impact a child’s achievement and progress. No matter how well a teacher plans and structures learning tasks, it is the teacher’s ability to make different decisions for different students at different times that informs the power and effectiveness of the instruction. The moment-to-moment instructional decisions teachers make based on their observations and analysis of children’s learning behaviors are significant. The teacher teaches the child, not the book or program.

3. METHODOLOGY

To explore the potential impact of classroom teachers on student LbE experience, we enrolled teachers and students in five different technology & engineering classrooms in our study following IRB approval and collection of consent documents. Specifically, this research was conducted as part of a larger *National Science Foundation* grant (Award: 2101235) project between Purdue University, Brigham Young University, the University of Georgia, the International Technology & Engineering Education Association, and the Dekalb County School District in the greater Atlanta, Georgia area (USA). Each of the enrolled teachers, and their students, were working in a district-level course, *Foundations of Technology*. While the classrooms, locations, and teachers were all varied by school, the overall course goals and LbE items compared by students were the same. Fully recognizing the presence of myriad confounding variables resulting from the different classrooms, locations, and teachers, we set out to better understand the ramifications of teacher differences on student experience through our research question using both quantitative and qualitative means.

3.1. *Quantitative: ACJ session rank and parameter statistics*

We created five ACJ sessions—each consisting of the same items for comparison—and enrolled consenting students in each of the participating classes in these sessions. While all items ($N = 50$) included in each session were identical, the items viewed by individual students were varied (because of the algorithm used by the ACJ software *RMCompare*). Further, teachers were given the freedom to craft individual holistic statements (i.e., the criterion provided to students with which to make judgments) and introduce and debrief the project however they deemed best. Students made between 5-10 comparisons of backpack images and, in each instance, selected the backpack design they thought best. Following all the judgments, the rank order (1-50) and parameter values for the items in each session were collected and compared using both Pearson and Spearman correlation coefficients for the parameter values and rank order, respectively. The resulting rank orders, parameter values, and correlation coefficients were used in the quantitative data analysis.

3.2. *Qualitative: Analysis of student comments made during LbE and teacher interviews following the session*

In addition to the five LbE sessions, and the accompanying quantitative data, we also collected student comments made while completing the LbE decisions. In each instance, students were prompted to justify their decision of one item over another. These comments were collected via the online ACJ platform *RMCompare*. Thematic coding techniques following recommendations from Saldaña (2013) were used to explore these comments in line with the research objective of exploring the potential for differences in student LbE experience based on different teachers. This process involved an initial review of the comments which was completed to identify potential themes. Multiple reviewers completed this step and themes were compared and refined following the review. After solidifying themes, a subsequent review of all comments was completed wherein thematic codes were applied to all student comparison comments.

These analyses of student comments were undertaken as a means of potentially triangulating the quantitative findings from the LbE sessions with the overall goal of investigating our stated research question: How are student LbE experiences similar, or different, as a result of classroom teacher differences?

4. RESULTS

Following the LbE sessions, and the collection of the associated data, the rank orders and parameter values for the items in each session were collected and conditioned prior to statistical analysis. Additionally, all thematic coding findings from the qualitative analysis of the student comments and teacher interviews were organized for sentiment and triangulation with quantitative findings.

4.1. Quantitative Findings.

Both a Spearman (rank) and a Pearson (parameter value) correlation were run for the results of each session (see Tables 1 and 2) and instances of statistical significance were noted at both the $p < .05$ and $p < .01$ levels.

Table 1.
Spearman Correlations for Rank Orders

| | T2 | T3 | T4 | T5 |
|----|-------|------|-------|-------|
| T1 | .38** | .29* | .18 | .29* |
| T2 | | .32* | .47** | .39** |
| T3 | | | .34* | .29* |
| T4 | | | | .16 |

* Correlation is significant at the 0.05 level (2-tailed)

** Correlation is significant at the 0.01 level (2-tailed)

Table 2.
Pearson Correlations for Parameter Values

| | T2 | T3 | T4 | T5 |
|----|-------|------|-------|-------|
| T1 | .37** | .27 | .16 | .28* |
| T2 | | .31* | .47** | .42** |
| T3 | | | .29* | .27* |
| T4 | | | | .12 |

* Correlation is significant at the 0.05 level (2-tailed)

** Correlation is significant at the 0.01 level (2-tailed)

Of note, the results from most of the correlations were statistically significant suggesting similar results from the LbE sessions run in different classrooms. This is not entirely surprising as similar course outcomes across classrooms should lead students to similar decisions in the comparative judgments between items.

More interesting to us were three instances of non-statistical significance in the correlations between the session results of several teachers (Teacher 1 and Teacher 3; Teacher 1 and Teacher 4; Teacher 4 and Teacher 5). Of these, the correlation between Teacher 1 and Teacher 3 approached significance (and was significant in the rank order) and no further investigation was conducted.

Building on the differences discovered through the lack of statistical significance in the LbE session rank and parameter value outcomes, the correlations between Teacher 1 and Teacher 4 as well as the correlation between Teacher 4 and Teacher 5 were investigated further. This exploration was conducted as part of the qualitative analysis reported in the next section.

4.2. Qualitative Findings

The process of manually coding qualitative data required the identification of recurring themes within interviews or other transcripts, followed by the systematic documentation of noteworthy discoveries. In the context of this qualitative research, this method of thematic coding was employed to analyze students' comments pertaining to their LbE decisions and revealed several similarities and differences between the student rationales for those students in classes taught by Teachers 1, 4, & 5.

4.2.1. Teacher 1

Teacher 1 is a male teacher with several years of experience teaching in the school, district, and with the course/curriculum. The criteria selected by Teacher 1 for students to use in making comparative decisions was: "Imagine you are redesigning a backpack. Which image is a more creative approach to research in the design process?"

Comments (n = 239) were mainly coded using the following categories: holding capacity (n=49/239), functionality (n=41/239), creativity (n=60/239), and comfort (n = 16/239). Specifically, Teacher 1's students had comments centered on both the functionality of the backpack designs and the form (creativity) in these designs. This is an especially interesting finding given the holistic statement chosen by Teacher 1 which focused squarely on "which image is a more creative approach..." Example student comments around functionality include:

"This would help because my phone could be dying and I could be on the go and I can charge my phone on the go."

"I chose option B because not many people have kids, but everyone can use a backpack at some point. Also, many people would problem like the design on option B a lot more."

"I chose this because even though the camping and hiking bookbag can be put to use really well, I feel like the stroller can be put to use more efficiently."

Additionally, other student comments centered on the creativity of the backpacks shown. For example, one student remarked:

"Option B is very creative because they created their own backpacks out of different materials. It is also most likely handmade."

Overall, a notable feature of the comments from Teacher 1's class was the specificity of answers; while many student comments indicated a feature of the backpack (that was either included or not), the students in Teacher 1's class did so with more detail than those in other classes.

4.2.2. Teacher 4

Teacher 4 is a female teacher with many years of experience—both in teaching the class and in the school/district. The holistic statement chosen by this teacher was: "Which carrying approach is most usable and why?"

Not surprisingly, the comments ($n = 301$) from her students were most inclined to designs that were practical and usable—in line with the stated holistic statement—and this was the most common code for their responses (74/301). Additionally, Teacher 4's students were the only ones to occasionally point out flaws in the option they did not choose (as opposed to simply indicating positive qualities of the option they chose), though these still largely centered on reasons related to functionality. Example comments from Teacher 4's students included:

"Option B is clearly better than A. This is because it is more complex and useful, you can even charge your phone."

"Option B is more usable because you could take it around wherever you need it. Unlike a carrier on a bike, you cannot take that everywhere."

"There are more compartments for different things to stay more organized and it takes less work trying to carry it or keep it on."

4.2.3. Teacher 5

Finally, Teacher 5 is an experienced female teacher with similar years of experience to Teachers 1 and 4. Teacher 5 used the same holistic statement as Teacher 1 ("Imagine you are redesigning a backpack. Which image is a more creative approach to research in the design process?") but the analysis revealed that Teacher 5's students placed a higher value on self-expression (e.g., form) over functionality and practicality. Comments from students in Teacher 5's class ($n = 207$) fell into categories of creativity ($n=45/207$), appearance ($n=20/207$), and uniqueness ($n=19/207$). Overall, Teacher 5's students did not shy away from preferring more innovative designs, and, of the three classes reviewed here, these students placed a much bigger emphasis on creativity and individuality. For example, some students commented:

"The lion is a very creative back pack as opposed to a normal backpack that is used often."

"I like the way the backpack looks, and its aesthetic value is above the other."

"It is a very different design than most backpacks and seems like it could hold more."

5. DISCUSSION

Our observations of LbE implementation across different classrooms suggested that teacher differences in style, implementation, and facilitation may impact the experience of students as they engage in LbE. However, existing research into this idea was lacking (Bartholomew & Jones, 2020); therefore, our intent in this research was to explore the potential differences in student LbE experience based on their different teachers. This was a natural extension of other work around ACJ (Bartholomew, Strimel, & Jackson, 2018; Bartholomew, Strimel, & Yoshikawa, 2019) and LbE (Bartholomew & Yauney, 2022; Bartholomew et al., 2020; Bartholomew, Ruesch, Hartell, & Strimel, 2020; Mentzer, Lee, & Bartholomew, 2021) which has demonstrated the potential for this approach to improve student performance but has not investigated the specific nuances of *how* LbE should be implemented to be most effective. Therefore, we engaged students across several different classrooms in the same LbE session and explored the potential similarities and differences in their experience (e.g., their judgments and their comments made while comparing items).

Importantly, we did not attempt to assess the “accuracy” of student judgments (e.g., how well aligned their rank orders are with those of professionals or an established rubric) in these settings or the final results of the rank order or parameter values from each session; rather, we have taken the ranks produced by students and investigated how they related to one another to examine the consistency across students enrolled in different classrooms. This exploratory research, while informative, also does not explain *why* such results were obtained and we readily admit the presence of a variety of external factors that likely influenced the findings (e.g., school, teacher, schedule, classroom, neighborhood, and a host of other differences). We do note that all participating students were enrolled in the same course in the same district in the same state in the United States. Further, the set of items viewed by the students was identical; however, as can be seen from both the quantitative and qualitative analyses performed, there were distinct differences in both the items selected by students (by class) and their rationale for selecting one item over another. Importantly, we did not dictate to the teachers the holistic statement (or judgment criteria) they should use with their students; this was left to the teachers to decide. While student comments for both Teacher 1 and Teacher 4 seemed to align well with the chosen teacher criteria for judgment seen in the holistic statement (either emphasizing form or function), students in Teacher 5’s class emphasized form even though the holistic statement was specifically centered on function.

While the differences in student comments made sense—and we hypothesized that potential differences would exist—it was both interesting and insightful for us, as researchers, to unearth these differences. While the quantitative analysis revealed a difference (i.e., instances of non-significant correlation between rank order results), this analysis did not reveal why such a difference existed. Later qualitative analysis showed an emphasis on either a) practicality/usefulness (a.k.a. function) or b) individuality/creativity (a.k.a. form). A difference that was often aligned with the specific holistic statement crafted by the teacher.

The criteria of both form and function appeared across all classrooms (e.g., see the high levels of reliability in rank orders produced between classes), but there were differences in the rank orders produced by the students with Teachers 1, 4, and 5. Specifically, the qualitative analysis showed

that while Teacher 1 and 5's students' comments were more focused on function, Teacher 4's students' comments emphasized form.

These differences align with those found in other design research which shows that the criterion for success is often divided between form and function (Khan, Pitts, & Williams, 2016). For example, Bartholomew, Reusch, Hartell, & Strimel, (2020) compared ACJ sessions completed by assessors in different countries and found that an emphasis on either form or function was one degree of separation in the results between countries. Interestingly, while all students/teachers in this project were in the same country, state, and district, the debate between form and function remained and the consensus produced in each classroom differed.

Additional research is needed to parse out the *why* behind these findings, as well as the *what now?* With respect to the inception of these judgments, it may be possible to uncover how different cultures created in each class have contributed to differing student mindsets in LbE. More closely reviewing the course-level objectives (e.g., how do form and function fit into the course-level objectives) may be relevant to understand how various types of judgment are expected in the course. Perhaps the teachers' educational and professional background and their design pedagogical knowledge also play a role here. Further interviews with the teacher about their intentions in these LbE sessions and when teaching design, in general, may reveal sources of influence in students' judgment.

Then, we wonder if students' judgment differences and rationale translate into later action (e.g., if they focus on form during LbE, do they maintain focus on form while designing?) and what other ramifications (academic or otherwise) their different experiences in LbE have. Further research recommendations here include longer observations of student design work or interviews, and a more detailed scripting of the LbE experience (e.g., the introduction made by teachers as well as the debrief following LbE).

Recognizing that our findings are specific in setting and narrow in scope (e.g., the students enrolled in this study, their classes, teachers, schools, and district), we nevertheless recognize potential applications for our findings to other settings. In addition to our recommendations around the need for further research into LbE and teacher impact, we recommend teachers 1) recognize and 2) be intentional about their own use of LbE in the classroom with attention to the wording of the holistic statement for comparative judgments. For example, if *form* is more important than *function* for a given design task, the teacher should emphasize and teach the principles of form prior to, and as a part of, both the LbE and overall design experiences. Further, the wording of the holistic judgment statement should intentionally align with the desired outcomes of the teacher. Alternatively, if both *form* and *function* are to be equally valued, teachers must recognize the ways they may, or may not, be influencing students' perceptions of these criteria—in both their classroom discussions and the provided rationale for comparisons.

Teachers interested in using LbE may consider ways different items (i.e., viewed by students) or holistic statements (i.e., judgment criteria for selecting between items) could impact the learning in their classroom. We note that design classrooms often employ gallery walks and other comparative activities (Rodenbaugh, 2015); these activities may be improved the use of targeted questioning around *form* or *function* with a specific emphasis on principles, ideas, and skills a

teacher wishes students to identify or hone in on. Additionally, classes outside of the Design and Technology area may similarly benefit from engaging students in LbE activities; especially if an intentional emphasis is placed on various aspects (e.g., form vs. function) of an assignment or task.

6. FUNDING

This work was made possible by funding from the National Science Foundation, Award: 2101235.

7. REFERENCES

- Bartholomew, S. R., Mentzer, N., Jones, M., Sherman, D., & Baniya, S. (2020). Learning by evaluating (LbE) through adaptive comparative judgment. *International Journal of Technology and Design Education*, 1-15.
- Bartholomew, S., & Yaune, J. (2022). The impact of differentiated stimulus materials in learning by evaluating. In *Pupils' Attitudes Towards Technology 39*.
- Bartholomew, S. R., & Yoshikawa-Ruesch, E. (2018). A systematic review of research around Adaptive Comparative Judgment (ACJ) in K-16 education. Council on Technology an Engineering Teacher Education: Research Monograph Series.
- Bartholomew, S. R., Strimel, G. J., Zhang, L., & Homan, J. (2018). Examining the potential of adaptive comparative judgment for elementary STEM design assessment. *The Journal of Technology Studies*, 44(2), 58-75.
- Bartholomew, S.R., Strimel, G.J. & Yoshikawa, E. Using adaptive comparative judgment for student formative feedback and learning during a middle school design project. *Int J Technol Des Educ* 29, 363–385 (2019). <https://doi.org/10.1007/s10798-018-9442-7>
- Bartholomew, S. R., Ruesch, E. Y., Hartell, E., & Strimel, G. J. (2020). Identifying design values across countries through adaptive comparative judgment. *International Journal of Technology and Design Education*, 30(2), 321-347.
- Bartholomew, S. R., Strimel, G. J., & Jackson, A. (2018). A Comparison of Traditional and Adaptive Comparative Judgment Assessment Techniques for Freshmen Engineering Design Projects. *International Journal of Engineering Education*, 34(1), 20–33.
- Fountas & Pinnell (n.d.). *Elevating Expertise: Why Teachers Are the Single Most Important Classroom Factor In a Child's Learning Achievement*. Retrieved on June 3, 2023 from: <https://fpblog.fountasandpinnell.com/elevating-expertise-why-teachers-are-the-single-most-important-classroom-factor-in-a-childs-learning-achievement>
- Hattie, J. (2015). What works best in education: The politics of collaborative expertise. British Columbia Teachers' Federation.
- Hattie, J. (2017). Hattie Ranking: 252 Influences And Effect Sizes Related To Student Achievement

- Khan, T., Pitts, M., & Williams, M. A. (2016). Cross-cultural differences in automotive HMI design: A comparative study between UK and Indian users' design preferences. *Journal of Usability Studies*, 11(2), 45–65.
- Mentzer, N., Lee, W., & Bartholomew, S. R. (2021). Examining the Validity of Adaptive Comparative Judgment for Peer Evaluation in a Design Thinking Course. In *Frontiers in Education* (p. 492). Frontiers.
- Pollitt, A. (2012). The method of adaptive comparative judgement. *Assessment in Education: principles, policy & practice*, 19(3), 281-300.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American economic review*, 94(2), 247-252.
- Rodenbaugh, D. W. (2015). Maximize a team-based learning gallery walk experience: herding cats is easier than you think. *Advances in Physiology Education*, 39(4), 411-413.
- Saldaña, J. (2013). *The Coding Manual for Qualitative Researchers* (Second). Los Angeles, CA: SAGE.

6. REFERENCES