PATT 39

ON THE EDGE

TECHNOLOGY, INNOVATION & EDUCATION

PROCEEDINGS

DESIGNING A BETTER *WORLD* THROUGH TECHNOLOGICAL LITERACY *FOR ALL*

June 21 – 24, 2022
Memorial University,
St. John's, NL, Canada

Editors:
David Gill, Jim Tuff, Thomas Kennedy, Shawn Pendergast and **Sana Jamil**

MEMORIAL UNIVERSITY

MARINE INSTITUTE

PATT

100 years of excellence
FACULTY OF EDUCATION
MEMORIAL UNIVERSITY

Newfoundland Labrador
CANADA

Newfoundland & Labrador English School District

NLTA

CONSEIL SCOLAIRE FRANCOPHONE PROVINCIAL

PATT 39

PATT on the Edge

Technology, Innovation and Education

St. John's, Newfoundland and Labrador, Canada

June 21st-24th, 2022

Edited by:

Dr. David Gill
Memorial University
Mr. Jim Tuff
Memorial University
Dr. Thomas Kennedy
Memorial University
Dr. Shawn Pendergast
Memorial University
Ms. Sana Jamil
Memorial University

The Impact of Differentiated Stimulus Materials in Learning by Evaluating

Scott R. Bartholomew[1*], Jessica Yauney [1]

[1]Technology & Engineering Studies, Brigham Young University: Provo, UT, United States

[*]Correspondence: scott.bartholomew@byu.edu

Abstract

Previous work has demonstrated the capacity for positively influencing student learning by engaging them in evaluation of previously submitted work as an intentional priming exercise. This exercise has been referred to as Learning by Evaluating (LbE). Nuances such as who should complete the evaluations, what should be evaluated, how often, and when, are all areas needing additional research and exploration. Expanding on current LbE research, we set forth to investigate the impact on student learning of intentionally differing the quality of examples evaluated by the students. In this research, university design students (N = 468) were assigned to one of three treatment groups; while each group evaluated previously collected student work as an LbE priming activity, the work evaluated by each group differed in quality. Using a three-group methodology, one group of students only evaluated high quality examples, the second only evaluated low quality examples, and the third group of students evaluated a mix of high- and low-quality examples. Following these LbE priming evaluations, students completed similar group projects that were then evaluated to determine if there was a difference between student achievement by treatment condition. Additional qualitative analysis was completed on student LbE rationales to explore similarities and differences in student behavior based on intervention grouping. No significant difference was found between the groups in terms of achievement. However, several differences in group judgement approach were identified and future areas needing investigation were highlighted.

Key Words: Adaptive Comparative Judgement, Learning by Evaluating, Assessment

1. INTRODUCTION

In general, assessment practices have improved over time (Robertson, Humphrey & Steele 2019), but relatively little has changed about students' participation in assessment processes, with assessment being viewed as a teacher-centric activity and having minimal effect on student learning (Johnson et al. 2019). Recent work with assessment and evaluation has demonstrated the potential of these activities to play a larger role in students' learning. Specifically, as students engage in evaluation activities and revisit/review/revise their own—or their peers'—work, their learning has been significantly positively impacted (Bartholomew, Mentzer, Jones, Sherman, & Baniya, 2022). Research into this approach, called LbE has highlighted this potential; instead of viewing assessment as a task meant solely for teachers, students intentionally engage in an evaluation process as a step in their own learning. LbE has demonstrated that as students engage with sample work, they exercise higher order thinking skills that can help strengthen their own understanding of the task, the associated requirements, and the applicable skills, aptitudes, and approaches (Bartholomew et al. 2020).

2. LITERATURE REVIEW

The theoretical basis for LbE is informed by multiple areas of study, including cognitive apprenticeship, and Bloom's taxonomy. However, despite these theoretical underpinnings, LbE is still in its infancy and the potential connections and implications warrant additional consideration.

2.1. Learning by Evaluating

Following research into improving design education, researchers recognized the potential for utilizing evaluation as a learning tool for students rather than simply as an assessment approach for teachers. Specifically, Bartholomew, Mentzer, Jones, & Sherman (2020) coined the acronym "LbE" to describe a process wherein students view and evaluate examples of work using Adaptive Comparative Judgement (ACJ) prior to engaging in similar assignments themselves. As a learning intervention, LbE starts an assignment with student evaluation of previously submitted work with the primary goal of informing their own future learning and work. Several studies have shown positive results in student learning through LbE with implications for facilitating student learning and growth (Baniya et al. 2019; Bartholomew, Mentzer, Jones, & Sherman, 2020; Bartholomew and Strimel 2019; Bartholomew et al. 2018b; Bartholomew et al. 2018a; Seery and Canty 2017). Students have called out benefits of this approach such as its ability to help them gain confidence (Canty 2012) and improve their own work (Bartholomew et al. 2019). This process has been shown to have positive effects in design, English, Engineering, and Business courses (see Bartholomew & Jones, 2020).

2.2. Cognitive Apprenticeship

In an LbE setting, a learner can critically evaluate previously submitted work and engage in several methods inherent to cognitive apprenticeship including reflection, modeling, and articulation (Collins, 2021). During the reflection portion, a student is then invited to compare their own thinking processes to the processes of experts or other students. Likewise, students engage in critical evaluations of their peer's work through pairwise ACJ comparisons and compare what is displayed (modelling) with their own ideas, thoughts, plans, and intentions for that assignment. Bartholomew, Strimel, & Yoshikawa (2018) highlight the benefits of this comparison process and highlight the opportunity for subsequent reflection (articulation). Given these connections between cognitive apprenticeship and LbE, we posit that cognitive apprenticeship may provide a rationale for, and a theoretical basis from which to build our understanding of LbE.

2.3. Bloom's Taxonomy of Learning

In addition to connections with cognitive apprenticeship, elements of LbE are tied to Bloom's Taxonomy of learning (1956). As students compare examples of previous student work and evaluate which is better, they practice two of the highest skills along Bloom's learning taxonomy (analysis and evaluation). Further, the act of articulating an evaluation decision rests firmly in the evaluation portion of Bloom's taxonomy. In this way, LbE rests on the premise that engaging students in higher-order thinking skills may lead to great learning gains (Collins, 2014) as opposed to other learning activities with an emphasis on lower-order methods.

2.4. Adaptive Comparative Judgement

Although LbE is not dependent on ACJ, the research into LbE has largely utilized ACJ as a vehicle through which the LbE comparisons are made (Bartholomew & Jones, 2020). By itself, ACJ is a method of comparing items with the goal of forming a rank order of item quality. In ACJ an individual views pairs of items and determines, based on an identified criterion, which is better. This process is repeated iteratively by multiple judges with the result of a rank-ordered list of all items determined by the collective comparative judgments. Previous research has shown high reliability levels (Baniya et al. 2019, Bramley

2015), a simpler assessment process (Kimbell, 2021), and greater ease of integrating assessment feedback from multiple assessors (Bartholomew and Yoshikawa 2018; Kimbell 2012b) as benefits of ACJ over other more traditional approaches (e.g., rubrics) to evaluation.

3. METHODOLOGY

With LbE research demonstrating the potential for enhancing student learning, questions around the potential to modify or enhance LbE have risen (Bartholomew et al. 2020; Buckley, Kimbell, & Seery, 2021). If LbE has demonstrated a significant impact on student learning using previous student work of varying quality, can LbE be improved even further through intentionally selected items for evaluation? In response to this call, we determined to investigate this potential. The research question which guided our efforts was:

RQ: What is the impact, if any, on student learning through LbE with differentiated stimulus materials?

Based on previous work (Bartholomew et al. 2020; Bartholomew et al., 2022; Kimbell & Seery, 2021), we hypothesized that varying the quality of examples during LbE would influence student learning differently. Specifically, we posited that students engaged in LbE with only high-quality examples would rise to a "higher standard" and have better educational outcomes - a hypothesis based on observations as educators who have used LbE, and which aligns with research that high expectations generally lead to increased student achievement (Johnston et al. 2019). We counter-hypothesized that mixed quality examples may lead to the highest educational outcomes based on an opportunity to identify strengths of high-quality examples and weaknesses of low-quality examples (e.g., research on contrasting cases by Miksza, 2011; work learning from others mistakes by Caniglia, 2020). Our specific hypotheses were:

H_0: Students who view only high-quality examples will perform better than their peers who only view low-quality examples during LbE.

H_1: Students who view mixed-quality examples will perform better than their peers who view only high-quality or low-quality examples during LbE.

To better understand the impact of example quality on student learning in LbE we used a mixed methods study with three collections of items: 1) high quality examples, 2) low quality examples, and 3) mixed quality examples. All examples were previously created student work centered on design point-of-view statements (POVs). Specifically, POVs identify a user, their unique need, and an insight for designing (see Wible, 2020; or Dam & Siang, 2020). The quality of each utilized POV was determined through an ACJ session conducted by course instructors in a previous section of the course following which the rank order was used to separate 125 student POVs into high-, low-, and mixed-quality groups. Specifically, POVs ranked 1-31 were categorized as "high-quality" examples, POVs ranked 95-125 were categorized as "low-quality" examples, and every fourth POV in the ranking (e.g., 1,5,9) were categorized as "mixed-quality" examples. The results from this process were 31 POVs for each treatment groups to evaluate during LbE.

Following the creation of our three POV sessions, we engaged 468 students in an introductory undergraduate design course at a large Midwestern University in our study. At the time of our intervention, these students were working in groups of 3-5 (N = 112) to complete an 8-week design project. Student groups were randomly assigned to one of three treatment conditions (high-, low-, or mixed-quality examples). Each section of the course included about nine teams with approximately three of these teams assigned to each condition. In this way, we attempted to mitigate teacher and/or section level variances. We intentionally opted to provide all students with the LbE experience—choosing not to include a traditional

"control group"—based on previous research which showed that LbE provided better educational outcomes for students (Bartholomew, Mentzer, Jones, & Sherman, 2020).

As part of this project, each group was assigned the task of creating a POV statement which would serve as a guide for their subsequent efforts. All students were introduced to the ACJ software to be used during class (*RM Compare*) and, as part of the assigned homework, students were provided with their assigned login and instructions for completing the assigned LbE. Students (N = 468) completed these LbE comparisons by viewing assigned examples, making comparative judgments on quality, and providing a rationale for each decision made. Each student viewed six comparisons (12 POVs)—differentiated in quality by treatment condition—and chose the better of the two using an online software (Figure 1). All evaluations, typed rationales, and other ACJ-generated data were collected and separated by treatment condition for analysis.

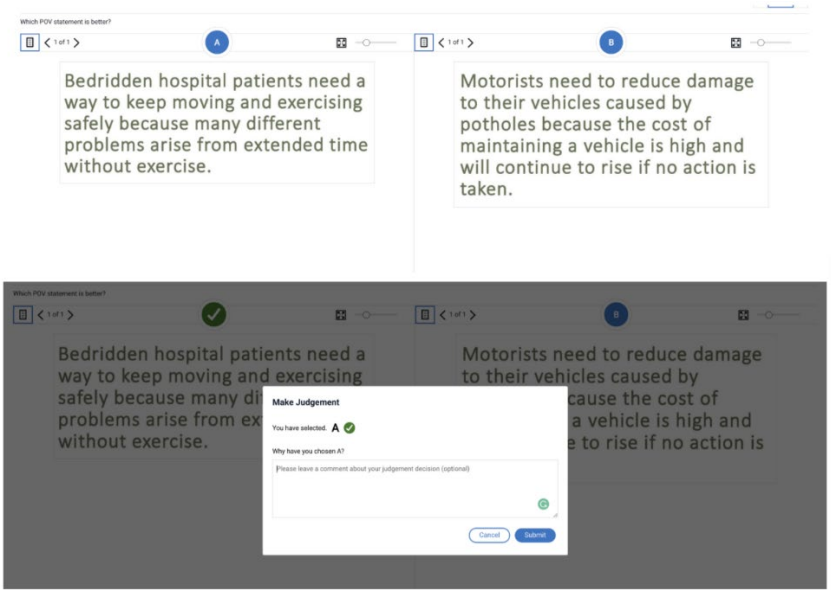

Figure 1. LbE POV student view

Following this intervention, all students worked in their groups during class to create POV statements and the remaining requirements of their project. At the conclusion, all student groups submitted their final POV statements for assessment; these final POVs (N=112), were then evaluated by students through an additional ACJ session to investigate the potential for differences in quality by treatment condition.

Our decision to utilize the students as evaluators during this second ACJ session was intentional - previous research (e.g., Bartholomew, et al., 2022; Strimel, Bartholomew, Purzer, Zhang, & Yoshikawa-Ruesch, 2020) has consistently demonstrated high levels of reliability between students' ACJ evaluations and the results produced through student ACJ sessions have demonstrated strong correlations with course instructors and industry professionals (Bartholomew & Jones, 2020). Given the large number of items (112 POVs), the burden of completing a significantly high round of judgments was alleviated by utilizing all enrolled students (N=468) as judges. The resulting rank, following 39 rounds of judgment, had a high reliability, $r = .83$.

The following data were collected, used, and analyzed in line with our stated research questions and hypothesis (see Table 1).

Table 1. Data collection approaches and statistical analysis

Data	Collection approach	Statistical Analysis & Associated Variables
High-Quality POVs Low-Quality POVs Mixed-Quality POVs	Collected from work submitted by previous students – ranked in an ACJ session by course instructors	
Treatment student (all 3 groups) LbE evaluation decisions Treatment student (all 3 groups) LbE evaluation rationale	Collected through the ACJ software *RMCompare*	Attribute Coding Frequency Counts Thematic Categories Pattern Coding
Treatment student (all 3 groups) POVs	Collected through the course learning management software	
Parameter Values (rank & magnitude) of treatment group student POVs Student evaluation rationale for treatment student POVs	Collected through the ACJ software *RMCompare*	ANOVA; Treatment group (IV) Parameter Values (DV) Pattern Coding

Statistical analyses were performed on the quantitative data using associated software (SPSS, V24) to determine what difference, if any, existed between the final student POVs and how the student learning was impacted by the quality of POVs viewed.

Additionally, three exploratory analyses were completed on students' comments collected during their LbE experience. Analysis of the prevalence of terms was used to explore potential trends regarding what students may be learning through LbE. This process involved an analysis of open ended-responses following suggestions by Feng & Behar-Horenstein (2019) and Saldaña (2015). All comments (2299) made by 402 students were analyzed using attribute coding with frequency counts. In this analysis the comments from student's evaluations during the LbE exercises were combined and the number of times relevant and related words were used in the decision rationales were identified. Data was sorted in line with the intervention groups and non-relevant words that did not contribute to the overall meaning of each comment (e.g., "the", "and") were removed. This frequency list was then sorted by frequency and used as a means of illuminating and triangulating findings derived from the other analyses.

Next, as part of the general inquiry into the potential for influencing student learning through intentionally varying the quality of items viewed during evaluations, we analyzed the overall sentiment of students' rationales using attribute coding. Each of the 2299 LbE comments was coded as either purely positive, purely negative, or neutral following Saldaña's (2015) recommendations using thematic categories. Student comments that provided positive feedback generally included words like "good," "better," and "more organized." These were coded as positive while student comments that provided negative feedback using words like "worse," "more confusing," and "missing" were coded as negative. All comments that included both positive and negative comments were coded as neutral.

A final qualitative analysis sought to identify if themes of students' comments provided during the LbE experience correlated with the themes of feedback they received on their own projects at the conclusion of

the POV creation process. 20 students were randomly selected from each treatment and a qualitative analysis of the six rationales they provided during the LbE experience was completed in line with recommendations from Baker & Edwards (2012). All elements referenced in the students' comments (both their rationale during the LbE intervention and the rationales provided by peers during the final POV evaluation) were identified. For example, if a student commented "It is a more concise problem that is actionable with more direct purpose" then the themes of being "concise" and "actionable" would be identified. Finally, each of these comments was qualitatively analyzed looking for potential correlation between the themes in students' LbE rationale and the rationale provided by peers during the final POV. Table 2 provides an example of our qualitative approach to coding.

Table 2. Example of analysis completed regarding student themes.

Initial Student Comments	-They started out with a long-winded POV statement but eventually came out with a refined version that I think fits very well. -I think option B is more **actionable** while maintaining **focus** and direction. Both statements have clear **stakeholders**, **needs**, and insights. -While A is a longer POV statement, I think it is still more defined and **focused**. -This one melds the components of a strong POV statement slightly better than the other. -I like the **insight** of this one. It is more surprising. -This one appeals to me. I want to see how it would work. The videos we watched in class explaining POV statements said that a good statement is **attractive** and option B is attractive to me.
Themes Identified	Actionable Focus Direction Components Insight Attractive Stakeholder Need
Feedback Received on Student's Project. Feedback about the project it was being compared to is crossed out and repeated themes are bolded.	I choose this one because although 'it is **short,** ~~the other one does not follow the format very well and explain why that is a problem~~ ~~B is too repetitive.~~ ~~Stakeholders are well defined, and the POV gives a very good explanation of what needs to be solved~~ ~~B is too long and not specific enough.~~ ~~The stakeholder is more descriptive than A's stakeholder.~~ ~~Option B is more focused and therefore gives the design team more direction in brainstorming and ideation.~~ ~~Again, simple states the problem without a broad solution~~ ~~This POV clearly has an actionable plan and provides insight for the design team.~~ ~~B provides better insights.~~ More **specific** to an issue and more **actionable** ~~This clearly identifies the stake holder.~~ This one is more **actionable** and has a better **stakeholder.** ~~Problem is more defined,~~ Option A is **too broad** and doesn't provide the same **insight.** The **stakeholder** is clearly defined ~~opposed to the less specific use of the word "people" in POV statement B~~ ~~It is more descriptive.~~ ~~The stakeholder, the need, the insight is clearly stated~~ ~~This statment is very developed and specific in each criteria.~~

	Option B is **ambiguous**, problem is not defined well enough for a solution to be created. The POV statement is more **detail**ed than B. I chose B because A went into more detail and had **too much** in its statement. It seems easier to **identify all the parts** needed in a good POV statement. Has all the **requirement components** in order to be a good POV statement. ~~This statment is more complete and specific~~ ~~A is more specific than B~~ PoV A seems to have a properly identified group of **stakeholders/users** compared to PoV B, as PoV B doesn't ~~specify who within the operations needs to cut down on their production / consumption.~~ ~~This is very clear and identifies all required areas.~~ ~~User is well defined, and need and insights are well explained~~ ~~The stakeholder, the need, the insight is clearly stated~~ It has a **user, a need,** and an **insight** with much fewer words than A. Very simple, **to the point** without too much detail Has all three descriptively. better **focus** ~~Less vague than the other.~~ This statement is clear and **consise.** POV B is too **vague** and does not offer a solution or any **insight** for the design team. ~~A provides more of a focus and direction for the design team.~~ ~~It is more descriptive.~~ B is **unclear** ~~A has better insights.~~
Correlation Between Themes	Students commented positively about the sample project and received positive feedback on their project regarding the same theme 13 times. Students commented positively about the sample project and received negative feedback on their project regarding the same theme six times.

4. RESULTS

Based on previous findings demonstrating that students who use LbE have better academic outcomes than those who do not (Bartholomew, Mentzer, Jones, & Sherman, 2020), we determined to investigate the potential, if any, to influence student learning outcomes by intentionally differing the quality of items evaluated during LbE. Using an ANOVA, our analysis revealed no statistically significant difference between students who were exposed to high, low, or mixed quality examples ($p = .809$). Specifically, the difference between high quality ($M = .13$, $SD = .836$), low quality ($M = -.07$, $SD = .970$) and mixed quality ($M = -.04$, $SD = 1.12$) groups was not significant either overall or between each of the Groups. Further, each group had a similar number of items ranked in the top and bottom quartiles.

We next investigated the potential for differences between the treatment groups in the qualitative data provided by students' comments on the six example POVs they were shown. The first analysis of open-ended comments consisted of analyzing the words included in the student rationale provided during the LbE intervention. This analysis of word frequency (see Table 3) revealed that the vocabulary specific to elements of POV statements were the most common with "need", "stakeholder", and "user," each appearing more than 250 times in students' comments. Following these words there were many instances of descriptors in the student feedback with terms such as "clear", "specific", and "detail" each appearing more than 150 times in students' comments. Using a chi-squared test, we noted that within these groups of commonly appearing words there was no significant difference between groups. This use of content specific vocabulary and critiques of writing style between groups supports our quantitative finding of no significant difference in the outcomes of student groups.

Table 3. Word Frequency in Student Comments by Group

Term	High-Quality POV	Mixed-Quality POV	Low-Quality POV	**Total**
need	163	195	195	553
clear	130	147	151	428
insight	139	153	127	419
stakeholder	118	146	120	384
specific	115	95	86	296
user	108	67	104	279
solution	70	81	103	254
focus	59	77	53	189
action	67	65	57	189
detail	69	59	43	171
format	37	53	58	148
define	33	47	39	119
point	16	39	42	97
revised	43	32	20	95
concise	12	44	38	94
descriptive	29	21	27	77
long	10	30	30	70
problem statement	17	25	23	65
real	22	16	22	60
thematic	16	25	13	54
vague	25	13	16	54
short	9	19	22	50

The second analysis of student comments involved categorizing comments as either positive, negative, or neutral. In all groups, there were more than twice as many positive comments as neutral comments and even fewer negative comments (See Table 4). Further analysis showed the group only exposed to high quality examples had significantly fewer negative comments than the other two groups - an intuitive finding given the high-quality nature of the items they compared. Overall, students were more likely to justify their judgement with positive comments than critical ones, regardless of treatment condition.

Table 4. Comment Sentiment by Group

	Positive	Negative	Neutral
High-Quality POV	483	95	165
Mixed-Quality POV	495	152	197
Low-Quality POV	443	140	207

The third qualitative analysis of student LbE rationale compared the comments provided by students during LbE to the feedback that they received on their own project. Specifically, all comments were coded as positive, negative, or neutral and the counts of comments were analyzed for potential differences (See Table 5).

Table 5. Instances of shared themes in student initial comments and feedback received

Group	Negative -> Positive	Positive -> Positive	Positive -> Negative	Negative -> Negative
Mixed-Quality POV	16	97	14	6
High-Quality POV	0	148	19	6
Low-Quality POV	5	146	37	1

Again, our analysis demonstrated no significant difference between groups. All groups received at least four times as many positive comments as negative comments during the POV evaluation.

While few differences existed between groups, several findings hinted at how students engaged in LbE. For example, students' LbE comments generally followed a theme centered on one aspect of a POV across all examples. For example, one student's feedback referenced "groups" 5 times and "stakeholders" 2 times. While user groups are an integral part of POV statement creation, they are just one part. This pattern of feedback around one aspect/idea was common across many student evaluations.

Another trend was a level of quality conditioning that appeared to impact student judgments. For example, students approached the judgment process relatively, meaning, they made judgements based on the caliber of examples they were seeing. Even students who were only exposed to high quality examples sometimes concluded that "Both of these are poor," and students who were only exposed to low quality examples concluded that "Both of these were very good."

5. DISCUSSION

Previous analyses clearly support the use of LbE in the classroom. This attempt to understand the types of examples that should be presented to students presents additional questions around the LbE process our analysis suggests that students may experience positive learning impacts regardless of the types of examples

given. Given this finding, researchers should continue to analyze other aspects of LbE such as the effect of prompts, teacher-led discussions, and peer review on student learning. Interviewing students and teachers about their experiences using LbE could provide greater insight into how teachers facilitate learning and what thought processes students participate in.

The fact that students exposed to only high-quality examples made negative comments and that those exposed to only low-quality examples made positive comments raises questions about how students' expectations are shaped by what they see. A potential avenue for future research includes exposing students to new examples after the initial evaluations to see how students' opinions differ based on the examples they are exposed to. Exploratory research that identifies different types of learning activities would allow further research into optimal implementations of LbE.

6. REFERENCES

Baker, S. E., & Edwards, R. (2012). How many qualitative interviews is enough? Expert voices and early career reflections on sampling and cases in qualitative research. *National Centre for Research Methods*, 1–42.

Baniya, S., Chesley, A., Mentzer, N., Bartholomew, S., Moon, C., & Sherman, D. (2019). Using adaptive comparative judgment in writing assessment: An investigation of reliability among interdisciplinary evaluators. *Journal of Technology Studies*, 45(2), 24–45.

Bartholomew, S. R. (2017). Assessing open-ended design problems. *Technology and Engineering Education Teacher*, 76(6), 13–17.

Bartholomew, S. R. (2019). Teacher led assessment of learning to student led assessment as learning. *Beyond multiple choice conference*, Washington DC.

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*. https://doi-org.erl.lib.byu.edu/10.1007/s10798-020-09639-1

Bartholomew, S. R., Strimel, G. J., Garcia Bravo, E., Zhang, L., & Yoshikawa, E. (2018). Formative feedback for improved student performance through adaptive comparative judgment. *Paper presented at the 125th ASEE conference*, Salt Lake City, Utah.

Bartholomew, S. R., Strimel, G. S., & Yoshikawa, E. (2018b). Using adaptive comparative judgment for student formative feedback and learning during a middle school open- ended design challenge. *International Journal of Technology and Design Education*, 29(2), 363–385.

Bartholomew, S. R., & Yoshikawa, E. (2018). A systematic review of research around Adaptive Comparative Judgment (ACJ) in K-16 education. *2018 CTETE Monograph Series*. https://doi.org/10.21061/ ctete-rms.v1.c.1.

Bartholomew, S. R., Zhang, L., Garcia Bravo, E., & Strimel, G. J. (2019). A tool for formative assessment and learning in a graphics design course: adaptive comparative judgement. *The Design Journal*, 22(1), 73–95.

Bramley, T. (2015). Investigating the reliability of Adaptive Comparative Judgment. *Cambridge Assessment*, Cambridge, 36.

Caniglia, J. (2020). Promoting self-reflection over re-teaching: Addressing students misconceptions with "my favorite no." JME (Journal of Mathematics Education), 5(2), 70-78.

Collins, R. (2014). Skills for the 21st Century: teaching higher-order thinking. Curriculum & Leadership Journal, 12(14).

Dam, R. F., & Siang, T. Y. (2020). *Stage 2 in the design thinking process: Define the problem and interpret the results*. The Interaction Design Foundation. Retrieved November 15, 2021, from https://www.interaction-design.org/literature/article/stage-2-in-the-design-thinking-process-define-the-problem-and-interpret-the-results

Feng, X., & Behar-Horenstein, L. (2019). Maximizing NVivo Utilities to Analyze Open-Ended Responses. *The Qualitative Report*, 24(3), 563+. https://link.gale.com/apps/doc/A582635094/AONE?u=byuprovo&sid=googleScholar&xid=ee3c4f4d

Johnson, C. C., Sondergeld, T. A., & Walton, J. B. (2019). A study of the implementation of formative assessment in three large urban districts. *American Educational Research Journal*, *56*(6), 2408–2438. https://doi.org/10.3102/0002831219842347

Johnston, O., Wildy, H., & Shand, J. (2019). A Decade of teacher expectations research 2008–2018: Historical Foundations, new developments, and future pathways. *Australian Journal of Education*, *63*(1), 44–73. https://doi.org/10.1177/0004944118824420

Kimbell, R. (2007). E-assessment in project e-scape. *Design & Technology Education: An International Journal*, 12(2), 66–76.

Kimbell, R. (2012a). Evolving project e-scape for national assessment. *International Journal of Technology and Design Education*, 22(2), 135–155.

Kimbell, R. (2012b). The origins and underpinning principles of e-scape. *International Journal of Technology and Design Education*, 22, 123–124.

Kimbell, R. (2021). Examining the reliability of adaptive comparative judgement (ACJ) as an assessment tool in educational settings. *International Journal of Technology and Design Education*. https://doi.org/10.1007/s10798-021-09654-w

Miksza, P. (2011). A review of research on practicing: Summary and synthesis of the extant research with implications for a new theoretical orientation. *Bulletin of the Council for Research in Music Education*, (190), 51–92. https://doi.org/10.5406/bulcouresmusedu.190.0051

Pollitt, A. (2004). *Let's stop marking exams.* Retrieved from http://www.cambridgeassessment.org.uk/images/109719-let-s-stop-marking-exams.pdf

Pollitt, A. (2012). The method of adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 19(3), 281–300.

Pollitt, A. (2015). On 'Reliability' bias in ACJ. *Cambridge Exam Research.* Retrieved 23, April 2020 from https://www.researchgate.net/publication/283318012_On_'Reliability'_bias_in_ACJ

Rangel-Smith, C., & Lynch, D. (2018). Addressing the issue of bias in the measurement of reliability in the method of Adaptive Comparative Judgment. *In 36th pupils' attitudes towards technology conference*, Athlone, Ireland, pp. 378–387.

Robertson, S., Humphrey, S., & Steele, J. (2019). Using technology tools for formative assessments. *The Journal of Educators Online*, *16*. https://doi.org/10.9743/jeo.2019.16.2.11

Saldaña, J. (2015). *The coding manual for qualitative Researchers*. Thousand Oaks, CA: SAGE.

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of educational psychology*, 103(4), 759

Seery, N., & Canty, D (2017). Assessment and learning: The proximal and distal effects of comparative judgment. *Handbook of Technology Education*, Springer. https://doi.org/10.1007/978-3-319-38889-2_54-1

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34, 273–286.

Wible, S. (2020). Using design thinking to teach creative problem solving in writing courses. *College Composition and Communication*, 71(3), 399-425. Retrieved from http://erl.lib.byu.edu/login/?url=https://www-proquest-com.erl.lib.byu.edu/scholarly-journals/using-design-thinking-teach-creative-problem/docview/2378096479/se-2?accountid=4488

PATT 39

ST. JOHN'S / NEWFOUNDLAND & LABRADOR / CANADA

MEMORIAL UNIVERSITY

MARINE INSTITUTE

PATT

100 years of excellence
FACULTY OF EDUCATION
MEMORIAL UNIVERSITY

Newfoundland Labrador
CANADA

Newfoundland & Labrador English School District

NTA

CONSEIL SCOLAIRE FRANCOPHONE PROVINCIAL