

# Adaptive Comparative Judgment and Psychological Safety in Accounting Education

Scott R. Bartholomew

Brigham Young University, United States

Matthew D. Jones

Purdue University, United States

Steven R. Hawkins

Southern Utah University, United States

Jeff Orton

Southern Utah University, United States

Received: September 23, 2021 Accepted: October 25, 2021 Published: October 28, 2021

doi:10.5296/jet.v9i1.19046

URL: <https://doi.org/10.5296/jet.v9i1.19046>

## Abstract

Traditional Accounting curricula include instruction on the preparation of audit documentation, however, experiential instruction for students on the process of review-both as a reviewer and reviewee-is often scarce or missing. This study investigated a classroom intervention that engaged undergraduate students in peer-review activities to gauge how peer review and feedback impacted student performance and their perceptions of being able to engage in interpersonal risks. Using a case-method approach, students developed audit workpapers that were later peer-reviewed through a digital system utilizing adaptive comparative judgment (ACJ). Students' achievement scores were collected, and students also completed a pre/post survey on psychological safety. Our results indicate that student psychological safety increased over the course of the semester; however, the peer review process through ACJ did not significantly improve student performance within the class. The students responded positively to the intervention as an engaging learning process and effective in teaching real-world skills. Thus, this intervention provides an example of how

peer review activities could enhance the learning experience for students.

**Keywords:** adaptive comparative judgment, peer feedback, psychological safety, auditing, peer review

## **1. Introduction**

### *1.1 The Accounting Review Process*

Auditing Standards require that auditors maintain sufficient documentation to support their conclusions. “Among other things, audit documentation includes records of the planning and performance of the work, the procedures performed, evidence obtained, and conclusions reached by the auditor” (PCAOB 2016, AS 1251.02). Standards also require that all audit documentation be reviewed by at least one other audit team member (PCAOB 2016, AS 1201.05). The initial creation of audit documentation (Payne and Ramsay, 2008; Shankar and Tan, 2006; Andiola et al., 2018), as well as the review process (Fargher et al., 2005; Lambert and Agoglia, 2011; Brazel et al 2004; Frank and Hoffman, 2015; Agoglia et al 2010; Hun-Tong Tan and Trotman, 2003; Rosman et al., 2007; Payne et al., 2010; Harding, 2010; Bamber and Ramsay, 2000), are critical in meeting audit standards and completing a quality audit. In the case of failed audits, the finalized audit documentation is the key evidence in auditor negligence trials (Backof, 2015). This audit documentation is prepared by all members of the audit team, and some team members with limited experience will be given significant reviewing responsibilities. In fact, after only 5 years of experience, auditors may spend as much as 50% of their time reviewing audit workpapers prepared by other audit team members (Bamber and Bylinski, 1987; Asare and McDaniel, 1996; see also Sweeney, Suh, Dalton, & Meljem, 2017). Therefore, recruiting incoming auditors with review experience could not only be a competitive advantage, but insurance against forms of auditor negligence.

Recruiters report seeking students who have both technical and behavioral knowledge and skills (Rynes et al., 2003). While technical skills are obviously needed to complete and review audits, the behavior knowledge and skills are also crucial to ensure healthy team and intergroup dynamics (Plant, Barac, & Sarens, 2019; Piper, 2017; Chambers & McDonald, 2013). Not only must auditors be able to work closely with clients to perform and document audit procedures, but they must be able to give and receive direct and candid feedback about audit work and documentation that has been completed. In some cases, auditors will go through peer review performance evaluations several times a week, making the ability to foster psychological safety-a team environment where it is safe to take interpersonal risks, such as speaking up on issues or conflicts (Edmondson, 1999)-a critical skill for a successful team working environments (Rozovsky, 2015). The need for educating auditing students in the technical skills of review with additional practice in fostering psychological safety in teams is key for a healthy workforce.

Andilla et al. (2018) noted that accounting curricula include instruction on documentation preparation, but the instruction on the review process-both as a reviewer and reviewee-is often sparse or absent. Consistent with this, a recent study (Ulrich and Blouch, 2018) asked accounting professionals responsible for evaluating new hires to rank 63 auditing curriculum

topics based on importance to the profession and effectiveness of university instruction. Audit documentation was ranked first in importance, but only 43 in effectiveness, suggesting a major need for improvement in how universities teach accounting students to prepare and review audit documentation. This study investigated one classroom intervention aimed at closing this importance/effectiveness gap by engaging students in peer-review activities as an intentional learning activity. Additionally, we investigated the impact of these peer-review activities on the reported psychological safety among the students.

### *1.2 Learning Through Peer Critique*

Accounting, management, and business literature related to learning through review, and more specifically “peer review” or “peer critique,” traditionally center on feedback opportunities for peers. However, what is not always clear is whether these evaluation practices refer to the evaluation of individual attributes and traits (Greenberg & Liu, 2017) (i.e., peer evaluation of the work ethic of a teammate) or an evaluation of peers’ work (Bartholomew, Garcia-Bravo, Zhang, & Strimel, 2019) (i.e., peer evaluation of a portfolio or dossier produced by a colleague). While these differences may sometimes be subtle, we believe it is important to clearly identify *what* is evaluated during peer review scenarios - especially when referencing “peer review.”

Research in “peer review” and “peer critique” is dominated by investigations into ratings of peer traits; these evaluations have, at times, been linked with improved performance (Brutus & Donia, 2017; Stewart, Courtwright, & Barrick, 2017) while also being linked with minimal impacts on subsequent performance at other times (i.e. Drexler, Beehr, & Stetz, 2001; Kidwell & Bennett, 1993; Stewart, Courtwright, & Barrick, 2009), and even negative impacts on future performance (DeNisi, Randolph, & Blencoe, 1983). A host of other variables such as the size of social networks (Greenberg & Liu, 2017), comfort in reviewing and comfort with the review procedure (Brutus & Donia, 2017), perceptions of previously performed reviews (Fini, Jourdan, & Perkmann, 2018), and individual extra/introversion qualities (Erez, Schilpzand, Leavitt, Woolum, & Judge, 2014; Stewart, Courtwright, & Barrick, 2017) all factor into the influence of peer review on subsequent individual traits such as effort and trait improvement.

Conversely, “peer feedback” (when peers provide feedback on one another’s work) is an area with less research and understanding (Harari, Lemoine, Parsons, & Sharp, 2017). Research in this area, which is sometimes bundled with an evaluation of peer traits, has been linked with improvements in subsequent performance (Donia, O’Neal, & Brutus, 2015; Potter et al., 2017); for example, some research has found improvements in team member performance, confidence, and work outcomes through peer review - especially over time (Donia, O’Neal, & Brutus, 2015). The importance of prior experience with review and feedback (Domingo, Petkova, & Lamm, 2017), the review environment and cultural values of individuals-both those providing and receiving the feedback-have also been highlighted as important in the efficacy of the review process (Cubrich, Elicker, De Luque, Gabel-Shemueli, 2019).

However, despite these positive findings, not all peer feedback research has demonstrated such positive results; other efforts (Mayo, Kakarika, Pastor, & Brutus, 2012) have highlighted

potential pitfalls as peer feedback resulted in a deflation of personal perceptions following highlighted deficiencies (Kim & Kim, 2020; see also DeNisi, Randolph, & Blencoe, 1983). Further, peer feedback scenarios have also been linked to struggles around the credibility of feedback due to potential discrepancies (Bartholomew et al., 2019; Hamer, Purchase, Luxton-Reilly, & Denny, 2015; McCarthy, 2017; Zhang, 2019). Additional struggles related to the time, effort, and planning required for incorporating the feedback received in learning settings often inhibit widespread adoption (Black & Wiliam, 2018).

Importantly, this act of engaging in peer review and feedback can become more than simply an evaluation and scoring process - this act can become a learning experience for the reviewer which has the potential to positively improve students' abilities by shaping their thinking *during the act of reviewing* (Bartholomew, Mentzer, et al., 2019; Czaja, & Cummings, 2010). As students engage formatively in peer review and feedback to/from their classmates and compare peer work, significant improvement has been found over other students without the same opportunities (Bartholomew, Mentzer, et al., 2019; Bartholomew, Strimel, & Yoshikawa, 2018; Seery & Canty, 2017). Studies have shown that as students participate in peer review and feedback processes, they:

1. perform better than students without this formative practice (Li & Gao, 2016; Li, Liu, & Steckelberg, 2010),
2. experience both improved feedback quality and better assignment scores (Bartholomew, Strimel, & Yoshikawa, 2019; Gielen & de Wever, 2015),
3. improve their critical thinking ability (Sluijsmans, Dochy, & Moerkerke, 1998),
4. develop a better understanding of class material (Stefani, 1994),
5. engage in a more analytical approach towards assignment criteria (Nicol, Thomson, & Breslin, 2014), and
6. demonstrate increased engagement (Jurado, 2011).

### *1.3 Adaptive Comparative Judgment*

Adaptive comparative judgment (ACJ) is an approach to evaluation which utilizes iterative comparisons between pairs of items (Pollitt, 2012). ACJ, although originally designed as an assessment tool, has increasingly been utilized as a mechanism for engaging students in review, critique, evaluation, and learning (Bartholomew, Mentzer, et al., 2019; Bartholomew, Strimel, & Yoshikawa, 2019; Seery & Canty, 2017). ACJ began from work by Thurstone (1927) and later Pollitt (2004); both posited that using comparative judgments to evaluate quality through holistic approaches, rather than rubric-centric methodologies, was a way for improving assessment techniques and reliability. Comparative judgment (CJ) involves an assessor-student, teacher, or professional-working to discern qualitative differences between items as opposed to subjectively evaluating them and assigning point values. In these settings the assessor does not tally points, rather they make holistic comparisons and choose which item, of those displayed, is better (Kimbell, 2007, 2012a; Pollitt, 2004). These comparisons are repeated until a rank order of all the items is produced (Bartholomew, 2017). The addition

of an algorithm which intentionally-pairs items based on previous judgments led to the creation of adaptive comparative judgment (ACJ); the intentional pairing is done to increase both efficiency and reliability of the results.

Concerns related to the validity of this algorithm have been raised (Bramley, 2015), debated (Pollitt, 2015), and addressed (Rangel-Smith & Lynch, 2018), with an ever-growing body of research demonstrating positive effects following ACJ implementation in classrooms across grade levels (Bartholomew & Yoshikawa, 2018; Bartholomew & Jones, 2021). Originally designed as an assessment tool focused on summative settings, the positive effects of utilizing ACJ as a formative peer review tool have been documented in recent work (Bartholomew, Mentzer, et al., 2019; Bartholomew, Strimel, & Yoshikawa, 2019). Both studies by Seery, Buckley, Delahunty, and Canty (2019) and Canty (2012) used ACJ in higher education and found that students react well to using ACJ for formative peer review as it focused on assessing the student work as a whole and facilitated improvement across the entire student body in a class. Bartholomew, Strimel, and Yoshikawa (2019) reported that students engaged in peer review through ACJ demonstrated learning gains as a result of four key activities: 1) their exposure to peer ideas, 2) the opportunity to engage in peer review through comparative judgments, 3) the process of providing peer formative feedback, and 4) the opportunity to receive peer formative feedback from their classmates who reviewed their projects.

The entirety of the research into CJ and ACJ, the outcomes, and the various implementations, is beyond the scope of this work; however, syntheses of work related to ACJ in educational settings (Bartholomew & Jones, 2021; Bartholomew & Yoshikawa, 2018) reported indications of promise for this approach to the assessment of open-ended problems and for student learning through formative peer review. Further, several researchers have included student opinions relative to embedding ACJ for peer review. Bartholomew, Zhang, Garcia-Bravo, and Strimel (2019) quoted students who had engaged in ACJ for formative review and feedback:

*I liked how I could view other [student's] works and see what they did right in order to improve myself. I thought this was helpful because it allowed me to read criticism that I wouldn't think my project would have (p. 89).*

*It was interesting to see the others' work and I found it helpful to get the feedback that I did. It gave me greater insights into what I could do to improve my work (p. 89).*

Additionally, Kimbell (2018), pointed out that as students review and evaluate work they also learn as they are required to not only recognize but also articulate what makes something good, better, or best.

*They could say that this ... and that ... and this ... and that ... are components of it. And maybe also that when you see this in association with that it's especially good. We know that, in design terms, the act of expression pushes ideas forward. So too with this [ACJ] discourse, the act of [comparison and evaluation] begins to crystallize the construct for them. It makes a vague and intangible construct into something a bit more substantive (p. 185).*

There is increasing evidence that peer review experiences—specifically those facilitated through ACJ—may lead to increases in student learning (Bartholomew & Jones, 2021). However, this approach has not been used equally in all disciplines (e.g., the majority of ACJ peer review research centers on Design and Technology education) and an understanding of student experiences, particularly in how comparative judgments impact psychological safety, during ACJ is not entirely clear.

#### *1.4 Psychological Safety*

In addition to the potential for students to demonstrate learning gains through peer review and feedback there is also anecdotal evidence that students engaging in this process may feel more comfortable with the overall review experience. For example, as students evaluate peer work, they may also inherently begin to recognize both the strengths and weaknesses of their peers' work and also those contained within their own work (Czaja & Cummings, 2010). Repeated peer review experiences may thus lead to an overall comfort level with both reviewing *and* being reviewed by others. We conjecture that this may be a subtle, but important, element of added benefit within peer review and feedback experiences.

Importantly, there may often be a notion that peer review and feedback processes, if not anonymous, leads to tension or other conflicts in teams, groups, or classrooms as individuals exhibit frustration over the critiques from their peers. Though adults, adolescents, and children can respond negatively to destructive or even critical feedback (Baron, 1988; Achterberg, van Duijvenvoorde, Bakermans-Kranenburg, & Crone, 2016; Morales, Vallorani, & Pérez-Edgar, 2019), most related studies have shown generally positive reactions from individuals (e.g., Bartholomew, Strimel, & Yoshikawa, 2019). Research in the field of psychological safety—defined as a shared belief, amongst individuals, that an individual or team is safe to engage in interpersonal risks (Edmondson, 1999; Newman, Donohue, & Eva, 2017)—suggests that the group dynamic may be the antecedent to how one reacts to the interpersonal risk of giving and receiving feedback.

Interpersonal risks can include speaking up on a number of topics or issues, showing or sharing interest in one another, engaging in constructive conflict, and experiencing failure in experiments or risks (Edmondson, 1999, 2004). When a team or organization fosters psychological safety, they create the group setting to be a place where members can speak up, critique, fail, and debate without degradation or shame. Rozovsky (2015) discovered that of the five main components of successful teams (psychological safety, dependability, structure and clarity, meaning of work, and impact of work), psychological safety “was far and away the most important of the five dynamics we found—it’s the underpinning of the other four” (par. 11). Psychological safety has more to do with intergroup dynamics than simply feeling comfortable.

Research on the subject has exponentially increased from its inception in the 1960’s (Schein and Bennis, 1965) and its resurgence in the 1990’s (Kahn, 1990; Edmondson, 1999, 2019). Psychological safety has now been described as a vital leadership responsibility that is a deciding factor for the ability for employees to contribute, collaborate, grow, and learn (Edmondson, 2019; Bergmann and Schaeppi, 2016; Duhigg, 2016). The construct has been

studied in various industries and types of organizations (Zhang, Fang, Wei, & Chen, 2010), namely healthcare (Edmondson, Higgins, Singer, & Weiner, 2016; Kessel, Kratzer, & Schultz, 2012) and in education (Edmondson, et al., 2016; Higgins, Ishimaru, Holcombe, & Fowler, 2012; Schepers, de Jong, Wetzels, de Ruyter, 2008).

Educational settings offer opportunities for students to share new, and at times, tentative ideas. In these settings students attempt new procedures and strategies and navigate unknown ideas and concepts – this can sometimes lead to failure of one kind or another (Bransford & Donovan, 2005; Byrnes, 1998). Such behaviors can be seen as risky, especially among peers. Such risks in one's educational experience, have been defined as *intellectual risk taking* (IRT) (Byrnes, 1998; Byrnes, Miller, & Schafer, 1999; Clifford, 1991; Beghetto, 2009). As education does not take place in a vacuum but is, more often than not, in the presence of peers, a certain level of psychological safety is needed to facilitate such important risk-taking behavior in learning.

Garvin, Edmondson, and Gino (2008) described the building blocks of creating a learning organization, and in support of Rozovsky's (2015) findings, the first building block is psychological safety. One cannot learn if one is not in an environment that promotes a certain level of acceptance of others' ideas and contributions. Cajiao and Burke (2016) studied instructional methods focused on social interaction with 246 business students in Colombia. The instructional method significantly-directly and indirectly-increased learning behaviors among students when mediated by psychological safety. Student social interaction, among students and instructors, is impacted by the level of psychological safety that is cultivated in the classroom or learning environment. Though, as Howorth, Smith, and Parkinson (2012) found, while engendering psychological safety in short amounts of time may be less effective than over longer periods (i.e., 10 months), in practice it is not impossible and has been achieved in various settings (Cajiao & Burke, 2015; Morrison et al., 2019; Oakley, Felder, Brent, & Elhadj, 2004). Cultivating psychological safety in the classroom may be as imperative as it is in successful work teams.

To teach skills of working in teams, instructors often use self- and peer-evaluations-something especially common in business education settings (Ohland, et al., 2012). We suggest that this may not only be a pedagogical technique, but an opportunity to practice the real-world skill of peer review and feedback. Based on the role psychological safety plays in successful work teams and in learning environments, it follows that this should similarly be part of the educational experience of peer review and feedback. De Stobbelier and Ashford's research (2014) supports this idea; in a sample of 224 employee-supervisor dyads, they found that interdependent task relationships moderated by psychological safety increased the tendency for employees to seek feedback from their peers. Psychologically safe environments can turn previously competing entities, like peers at work, into sources of learning. Such environments may not eliminate conflict but may make conflict more productive. Given the impact of psychological safety in motivating individuals to seek out feedback, we contend that further research is needed as to the influence the strategies by which leaders or instructors implement peer review and feedback processes into the teams they oversee have on the construct. ACJ is one such strategy for peer review and feedback,

and though it has only once been used in business education to assist in peer evaluation (Metzgar, 2016), how ACJ influences psychological safety within teams is greatly under-researched. Therefore, a particular interest for our study was the interaction and impact of peer review using ACJ on one's ability to take interpersonal risks.

### *1.5 Research Questions and Hypotheses*

The guiding research question for our study is “Does a student’s use of ACJ for peer review impact their performance on accounting case study projects and/or their psychological safety during group work?” Based on the research discussed above, we expect that as students use ACJ to engage in peer review at the conclusion of each project in an accounting course, 1) their technical abilities will improve, 2) they will learn to recognize “good” solutions, and 3) they will solidify their own understanding of important accounting principles. Therefore, we form the following hypotheses:

H1: Students who use ACJ to evaluate peers’ projects will show greater performance improvements on future projects than student’s who do not evaluate peer’s projects.

Further, research suggests that the inherent collaborative nature of ACJ may impact student psychological safety. This psychological safety is needed in everyday teamwork but becomes vital in the case of interpersonal risks such as speaking up and in giving and receiving feedback. ACJ provides an opportunity to be exposed to peer feedback in an anonymous virtual environment with binary (e.g., “which one is better?”) judgments that results in a rank order of student work (best-to-worst). It is our hypothesis that this method of peer review and feedback in the class will impact student psychological safety, specifically in that as students use ACJ, they will improve in psychological safety and willingness to both provide and receive feedback. We link this hypothesis to four key traits of ACJ (Bartholomew, Strimel, & Yoshikawa, 2019): 1) Exposure to their peers’ ideas, 2) Providing feedback to peers, 3) Receiving feedback from peers, and 4) Deciphering between gradations of quality in peer/self-solution design. Taken together, these ideas lead to our second hypothesis:

H2: Students who use ACJ to evaluate peers’ projects will demonstrate increased psychological safety.

## **2. Method**

This research took place at an AACSB accredited University in the Western U.S.A. with students enrolled in a senior level introductory Auditing course. Of the 32 total class members, 17 were female (53%) and 15 were male (47%). All students enrolled in this course were randomly paired in teams of two students each. Each team completed six auditing case studies designed to represent work they will typically encounter during their first year in the auditing profession.

Five of the case studies were completed using Microsoft Excel, and one case was completed using Microsoft Word (see Table 1 below). Prior to completing work, the student teams were randomly organized into two treatment groups, Group A and Group B, for the intervention designed to evaluate the cases. Group A was engaged in using ACJ to compare, rank and



provide feedback (the intervention) for cases 1-3, while Group B was the control group for the first three cases. Group B was then engaged in the intervention to evaluate cases 4-6, while Group A was not.

Before the first case, the professor held discussions with the class to ensure understanding of why the intervention and practice of peer review and feedback was important in the class, what skills it was designed to help them develop based on his experience in the profession and feedback from current professional employers, and finally, how these skills would help them in their future careers. Class discussions were also held after cases 1, 2, 4, 5 and 6 to help the students reflect on the purpose of the intervention activity, and to discern if the students felt like the activity was accomplishing these objectives.

Table 1. Description of case studies and the format of deliverable used in the intervention

Case	Name	Format	Description and skills
1	Materiality	Microsoft Excel	Students were provided an Excel workbook with instructions and were asked to calculate a materiality limit for three separate scenarios based on a method for calculating materiality commonly used in the profession.
2	Planning Analytics	Microsoft Excel	Students were provided an Excel workbook with instructions and were asked to perform vertical and horizontal analytics on the Balance Sheet and Income Statement. Then students were asked to analyze the calculations for trends and outliers above predetermined thresholds to draft tick marks designed to develop the audit strategy based on their analysis of risk indicated in the data.
3	Internal Control Findings Memo	Microsoft Word	Students were presented with three different scenarios and asked to rate the significance of the internal control deficiencies and draft a professional memo to the client discussing the findings and their recommendations to improve internal controls.
4	Cash Audit	Microsoft Excel	Students were provided an Excel workbook with instructions and supporting documents to audit cash accounts. They were asked to perform an audit of the bank reconciliations for each of the cash accounts of the client by analyzing supporting documentation, and to prepare a conclusion of their findings, along with proposed journal entries, to correct any errors discovered.
5	Accounts Receivable Confirmations Audit	Microsoft Excel	Students were provided an Excel workbook with instructions and supporting documents to audit accounts receivable confirmations. They were asked to analyze confirmation responses for accounts receivable, then calculate a total projected error based on the errors discovered in the sample and propose a journal entry to correct the errors.
6	Accounts Payable Confirmations Audit	Microsoft Excel	Students were provided an Excel workbook with instructions and supporting documents to audit accounts payable confirmations. They were asked to analyze confirmation responses for accounts payable, then calculate a total projected error based on the errors discovered in the sample and propose a journal entry to correct the errors.

During each intervention activity, students in the treatment group were individually shown a pair of audit workpapers through the ACJ software (*CompareAssess*). Students viewed these pairs of workpapers and selected the item (audit workpaper) they believed was “better”

between the two items displayed. This decision was to be guided by their understanding of the correct answer (e.g., from the instruction provided by the instructor) and the professional formatting standards included on the assignment rubrics. Students were also prompted to provide a brief comment as to why they chose the “better” item for each decision they made (see Figure 1).

Each student compared approximately 12 pairs of items per ACJ session and completed the ACJ comparisons in less than 15 minutes on average. Two accounting professors also independently completed the evaluation process as a secondary check to ensure the perceptions of quality by the students were appropriate and aligned with accepted standards. In the class period following ACJ sessions 2, 4, 5, and 6, the course professor showed the class the anonymized ranking statistics, as well as any outlier statistics, and a visual of the highest ranked case. The purpose of displaying the highest ranked case was to provide an additional visual example, beyond the suggested solution provided by the professor in their review, of what a good example of the audit workpaper in the context of each case should look like (see Appendix A).

In addition to the ranking activity described above, cases were also manually graded by the course professor as separate validity check on the findings obtained. Instructions and a grading rubric were provided to the students to help them prepare each case and the same rubric was provided to the students for the ACJ ranking sessions to assist in their comparison. Additionally, the cases were graded for points before the professor participated in the ACJ session and the overall ranking results from the ACJ software were completed in order to create independent comparison points and mitigate the potential for bias in grading. The resulting statistics derived from the ACJ sessions (not from student grades) were used to investigate the guiding research question related to the impact on student performance of the learning by evaluating intervention.

### *2.1 Measures*

The outcome of an ACJ session consists of several data points including: a rank order of all items with parameter values for the items, misfit statistics for judges and items, and comments collected in conjunction with each decision. Of note, the parameter values, which were used in this research, provide insight over the rank alone as they signify both the ranking and the magnitude of difference between items (Pollitt, 2012) – they thus become a useful tool for analysis, interpretation, and investigation.

Misfit statistics are valuable as they indicate the relative agreement between judges and the relative agreement on the placement of each item in the overall rank. A judge with a large misfit statistic consistently makes comparative decisions contrary to those of their peers. An item with a large significant misfit statistic represents a “controversial” item—one that is ranked highly by some judges and very low by others. These misfit statistics were used in this research to investigate relative consistency among the student judges.

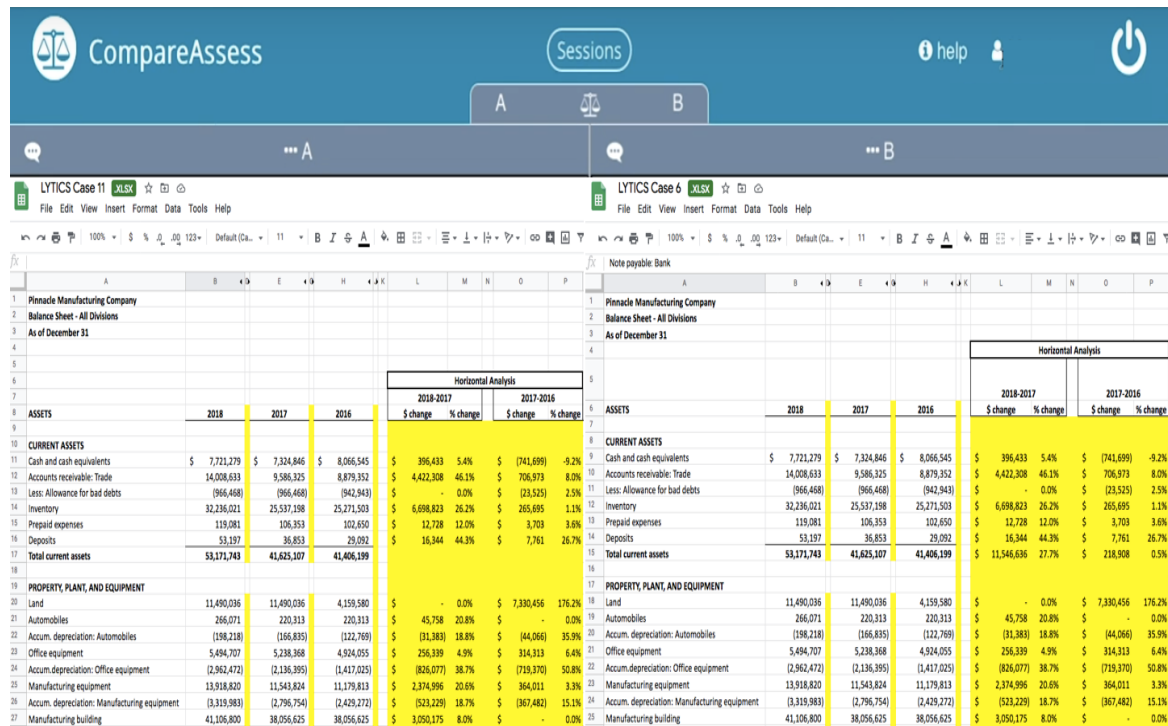


Figure 1. Visual layout of a CompareAssess ACJ session with two examples of student work side by side

Beyond the ACJ statistics, we created a pre/post survey based on Edmondson's (1999) Team Psychological Safety scale (Edmondson, 1999, 2019; see Appendix B). The 7 items were left unchanged in wording. All 7 items were based on a 5-point Likert scale ranging from strongly disagree to strongly agree. In addition to the survey, additional items were included in the post survey to collect student reflections on the use of ACJ in their learning experience over the course of the semester. In total, there were 7 items focused on psychological safety in the pre- and post-surveys with an additional 6 items aimed at the learning experience with ACJ in the post survey. We also collected personal information regarding gender and year in school but did not have participants include any other identifying information such as name or email, thus making the reporting as anonymous as possible for both pre- and post-surveys. These anonymity measures, guided by our IRB approval, created a limitation in our statistical analysis in that we could not perform a simple repeated measures analysis. Therefore, a more complex statistical method-a Cumulative Link Mixed Model-was needed to understand the average change(s) in responses over time.

The surveys were administered via Google Forms and students were instructed to take the assessment outside of class by the instructor. Points were given for those students who reported completing the surveys. The pre-survey was administered at the beginning of the semester. The case studies began during week four of the sixteen-week semester. The post survey was administered after case 6, which was during week eleven of the semester.

### 3. Results

#### 3.1 Differences in Parameter Values Across Groups and Time

To test our first hypothesis, we developed a mixed effects model with repeated measures on student groups comparing parameter values for the ACJ-participating and non-participating groups. We created the main effect of Case (signifying which case study was being performed) and another main effect of ACJ Group (signifying which group of students was participating in the ACJ for that particular case). In our study the main effects of ACJ Group and Case were not meaningful effects on their own; the effect of interest was their interaction. Lastly, a Kenward-Rogers approximation was used for testing the main effects in the mixed effects model we developed.

The main effect of Case was not statistically significant  $F(5, 32.1) = 0.17, p = 0.97$ ; neither was the main effect of ACJ Group,  $F(1, 16.7) = 0.53, p = 0.48$ . The interaction of these two factors was also not significant  $F(5, 32.1) = 0.57, p = 0.72$ . Ultimately, there was no statistically significant difference in performance between groups and their use of ACJ in the course thus leading us to reject our first hypothesis.

Table 2. Resulting statistics from student parameter values

	Df	Sum Sq	Mean Sq	F value	p value
ACJ.group	1	1.159166	1.159166	0.43058	0.52315
Case	5	0.567033	0.113407	0.095537	0.992593
ACJ.group:Case	5	1.960255	0.392051	0.330276	0.892878
Residuals	69	81.9058	1.187041	NA	NA

**Note:** ACJ.group describes which group of students participated in ACJ for the particular case; Case signifies which of the six case studies was being performed; and ACJ.group:Case represents the interaction effect.

#### 3.2 Differences in Psychological Safety Ratings Over Time

To test our second hypothesis, quantitative data was extracted from survey items 1-7 on both pre- and post-surveys through an analysis of the 5-point Likert scale. A basic two-way ANOVA looking at gender –being the main classification for subjects without other identifiers present– and test (pre- or post) as independent variables with the score as our dependent variable. Results showed an interaction between male and female responses from pre- to post, but this interaction was not statistically significant; males and females did not

have significantly different average scores (on average between pre- and post). However, on average, scores increased across genders from pre- to post. These increases are generalized to the group as student responses were not paired from pre- to post so group responses were combined.

A Cumulative Link Mixed Model (CLMM) was used to investigate the impact of individual items and their average change over time. Though it is not as powerful as that of the ANOVA, this model allowed us to have more resolution for the data set we collected. For example, items 2,4,5,6,7 were affected the least from pre- to post. Conversely, items 1 (*If you make a mistake on this team, it is often held against you*) and 3 (*People on this team sometimes reject others for being different*) were affected the most and our analysis revealed that psychological safety scores on items 1 and 3 were significantly improved for all students from pre- to post. This held true regardless of gender with a higher postsurvey value than presurvey for all participants. Additionally, further investigation revealed that, on average, there were greater odds of males giving a higher score on item 5 (*It is difficult to ask other members of this team for help*) on the postsurvey relative to the presurvey than females.

Table 3. Total psychological safety scores for pre- and post-surveys and averages for total scores and gender

	<b>Item 1</b>	<b>Item 2</b>	<b>Item 3</b>	<b>Item 4</b>	<b>Item 5</b>	<b>Item 6</b>	<b>Item 7</b>	<b>Total Avg.</b>	<b>Female Avg.</b>	<b>Male Avg.</b>	<b>Highest Possible Score</b>
<b>Pre</b>	2.79	3.27	3.31	3.31	3	3.79	3.55	19.482	18.94	20.25	35
<b>Post</b>	4.307	3.96	4.57	3.88	4.11	3.307	4.038	29.192	29.66	28.54	35

### 3.3 Student Comments on ACJ Experience

On the post survey, an additional six items (Items 8-13) were included to have students provide feedback and reflect on their experience in using Adaptive Comparative Judgment. Each student response was coded holistically (e.g., if a student had several sentences in agreement with ACJ as an effective learning tool, the comment was coded as “Yes” or “Like”) and by sentence (e.g., students provided reasons for why ACJ was an effective learning tool, and each reason or sentence was marked and tallied). In the following report of student responses, the term *Holistic Responses* refers to the entire answer provided to the question whereas *Evidence Statements* refers to mentions or sentences within the student responses. A student’s single response often made many statements (see Table 4 for an example of how student responses were coded). Table 5 displays the item questions, total counts among holistic response and evidence statement codes, and provides a connected student comment to provide context for the coding.

Table 4. Examples of how student qualitative survey responses were coded

**Item 8: How would you describe your experience using CompareAssess? Why?**

<b>Holistic Response (HR)</b>	<b>HR Code</b>	<b>Evidence-Statements (ES)</b>	<b>ES Codes</b>
<i>I thought it was fairly smooth and easy, however I feel the user interface could've been a little more user friendly, especially when looking at the criteria by which to judge the cases.</i>	Both (positive & negative)	<i>I thought it was fairly smooth and easy...</i>	Easy (Positive)
		<i>...however I feel the user interface could've been a little more user friendly, especially when looking at the criteria by which to judge the cases.</i>	Difficult program interface (Negative)

**Item 9: What did you like and not like about CompareAssess judgments?**

<b>Holistic Response (HR)</b>	<b>HR Code</b>	<b>Evidence-Statements (ES)</b>	<b>ES Codes</b>
<i>At first I didn't think that it was super clear what needed to be done and I didn't really like that I needed to write an explanation of why I was choosing my judgement. Those are the two things that I didn't really like about it. But one thing that I did like is that you could choose to look closer at just one if you wanted to and adjust sizing etc.</i>	Both (like & dislike)	<i>At first I didn't think that it was super clear what needed to be done...</i>	Confusing rubric/instructions (Dislike)
		<i>...I didn't really like that I needed to write an explanation of why I was choosing my judgement.</i>	Comments (Dislike)
		<i>...one thing that I did like is that you could choose to look closer at just one if you wanted to and adjust sizing etc.</i>	Program interface (Like)

The total number of students that voluntarily completed the final survey was twenty-six (out of 32 total students), therefore all the following descriptive statistics, unless otherwise noted, are derived from that result. Percentages were rounded to the nearest whole number.

*Item 8: How would you describe your experience using CompareAssess? Why?.* There were

three holistic response codes: *positive*, *negative*, or *both*. About 46% (12 of 26) of student responses were coded as *positive*; with nearly 15% coded as *negative*. The holistic response code, *both*, accounted for around 38% of the total responses. Two student comments were coded holistically (*positive*, *negative*, or *both*), but did not offer any further evidence statements (they did not answer the question “why”). Nearly half (46.15%) of the negative evidence statements from student comments focused on the difficulty of the program interface while nearly 38% of the positive evidence statements from student comments described how “easy” the program was to use.

*Item 9: What did you like and not like about CompareAssess judgments?* There were four holistic response codes: *like*, *dislike*, *both like and dislike*, and *no indication*. Student comments that were coded as *both like and dislike* contained evidence statements of like and dislike, whereas student comments holistically coded as *like*, for example, only accounted for reasons they liked the intervention. Of total student comments, over 57% (15 of 26) of them were holistically coded as *dislike*. Again, 15 of the 22 (about 68%) evidence statements for why the students did not like their experience were due to the program interface. Holistic responses of *like* and *both like and dislike* accounted for about 19% each while one student made no indication as to what they liked or did not like.

*Item 10: Do you feel that CompareAssess was an effective learning tool? Why or why not?* There were three holistic response codes: *yes*, *no*, and *maybe*. These three parent codes answered the first part of the question, whereas the evidence statements were used to answer the second (i.e., why or why not?). Nearly 69% of the student comments were coded holistically as *yes* comments. About 19% were *no* and just over 11% were coded as *maybe*. The total evidence statements for *yes* were 21 whereas only 5 evidence statements were associated with the *no* holistic response code. A large number (8 of 21; 38%) of the *yes*-associated evidence statements focused on the fact that the CompareAssess intervention was a “glimpse of real life.” Others (5 of 21; about 24%) found value in the act of comparing.

*Item 11: Do you see any potential value in using comparative judgment practices in your future career? Why or why not?* Much like Item 10, there were three holistic codes of *yes*, *no*, and *maybe*. The majority of students (65%) were coded holistically as *yes*, in that they did see potential value in using comparative judgment in their career. However, just over 23% of student comments were coded as *maybe*. Much of the associated *maybe* evidence statements highlighted that the comparative judgment’s use in a future career depended on the career, the setting, and other factors. Such variance demonstrates that it was not distinctly *yes* or *no* for these students. Nearly 65% of the *yes* associated evidence statements mentioned “skill building” as why comparative judgment had value to them.

*Item 12: As you made judgments, did you feel you learned anything? What specifically?* The three holistic response codes were *yes*, *no*, and *maybe*. The majority of student comments (over 69%) were holistically coded as *yes* with 23% were coded as *no*. There were 18 total evidence statements associated with the *yes* holistic response code. In half of those evidence statements, students reported that they had learned aesthetics (i.e., how to make a report “look good” or presentable) while only 3 students specifically mentioned the material or subject

(under 17%). A large number of students provided no evidence statements (nearly 67%) for why they did not feel they learned anything as they made judgments. Around 33% stated that they already knew the material before doing comparative judgment and thus, did not learn anything new.

*Item 13: What was the basis for your judgment decisions? What formed the criteria for your decision to choose one item over another?.* There were six holistic codes: accuracy, appearance, accuracy and appearance, self-reference, other, and none. Nearly half (46%) of the students reported that they looked for both accuracy-meaning, the documentation was completed correctly-and appearance-meaning, the documentation was well formatted. Over 15% of students only judged on appearance while another 23% only judged on accuracy. Interestingly, almost 8% of students reported using their own documents as a reference for their judgment decisions (i.e., “Was it better than mine” or “Did they get the answers I got”, etc.).

Table 5. Student survey responses: total counts, parent-child code breakdown, and student examples

Item	Parent Codes	Evidence-Statement Codes (Total)	Example
Item 8: How would you describe your experience using CompareAssess? Why?	Positive (12)	Easy (5)	<i>It was easy to pick and choose and all the materials were included.</i>
		Similarity Comparing (3) in	<i>It was interesting to have to compare audit reports that look very similar.</i>
		Self-Improvement (3)	<i>Good to see the different way others presented it and it helped me identify how to make minor adjustments to make it look more professional.</i>
		Hands On (2)	<i>Great, hands on experience.</i>
	Negative (4)	Difficult Program Interface (6)	<i>It was difficult to manage and view files on a single screen.</i>
		Similarity Comparing (1) in	<i>since all of our work papers were the same format I felt like it really didn't matter which one I picked.</i>
		Confusing rubric/Instructions (3)	<i>I think the rubric could be made easier to follow</i>
		Long (2)	<i>I felt like some of them were long and I just tried to get it done.</i>
		Lacked Training (1)	<i>I wish I had a bit more training on how to compare reports.</i>
	Both (10)	NA	<i>I thought that it was useful, but I think the rubric could be made easier to follow.</i>



Item 9: What did you like <b>and</b> not like about CompareAssess judgments?	Like (5)	Comments (2)	<i>I liked that I could leave a brief explanation...</i>
		Program Interface (3)	<i>I liked how smooth the system was, it made it very easy to actually compare and assess.</i>
		Comparing (3)	<i>I liked how quick we were able to compare both working papers.</i>
		Real life experience (1)	<i>I thought it worked really good for comparison to actually working in a corporate office.</i>
	Dislike (15)	Confusing rubric/Instructions (3)	<i>I was unsure what i was looking for on the first couple of assignments. For me it would have been more effective to have had more understanding of the paper work.</i>
		Program Interface (15)	<i>I didn't like that you couldn't choose one paper to look at, you had to have the screen split. Making it bigger was difficult and honestly took more time than the judgements.</i>
		Comments (1)	<i>I didn't really like that I needed to write an explanation of why I was choosing my judgement.</i>
		Similarity (3)	<i>I didn't like that our work papers were all the same format. I felt like it really didn't matter which one I picked.</i>
	Both Like and Dislike (5)	NA	<i>I liked the idea of comparing the 2 spreadsheets but some of multiple tab ones were more difficult to transition through.</i>
	No indication (1)	NA	<i>none</i>
	Item 10: Do you feel that CompareAssess was an effective learning tool? Why or why not?	Yes (18)	Recognize (1)
Glimpse of Real Life (8)			<i>Yes it was, it gave us a glimpse of what we will see in th[e] field.</i>
Added to the class (3)			<i>I believe it was an effective learning tool because it allowed us to learn beyond the actual cases and practice actual auditor duties.</i>
Comparing (5)			<i>Yes, by placing 2 different cases together, it was easy to judge.</i>
Self-Improvement (2)			<i>I learned what details to add to make a work paper sound professional and</i>

			<i>stand out.</i>
		Easy (2)	<i>Yes. Easy to use.</i>
	No (5)	Program Interface (1)	<i>No, it made comparing the options much more difficult than looking at them separately.</i>
		Did not learn anything (1)	<i>I don't feel like I learned anything from it</i>
		Similarity (1)	<i>It didn't feel like there was anything different about the papers we were judging.</i>
		No value in comparison (1)	<i>Comparing two entries on a purely objective basis was not particularly enlightening or intellectually stimulating</i>
	Maybe (3)	Glimpse of real life (1)	<i>I feel that it was a very simple process. If that is how it is in real life scenarios then I think it was an effective learning tool. If it doesn't relate to real situations then I think that it was probably not useful.</i>
		Timing (1)	<i>Yes and no. I feel like I didn't really know what I should be looking for so it might be a good thing to use right when we are starting out in this beginning course and then again later when we know what we should be looking for a little more.</i>
		Comparing (1)	<i>Not as effective, but it was a great experience to see how others approach their answers and all.</i>

Item 11: Do you see any potential value in using comparative judgment practices in your future career? Why or why not?	Yes (17)	Recognize (3)	<i>Yes, because audit is being able to recognize certain clues, and [CompareAssess] helps build those skills.</i>
		Reinforce (3)	<i>I think it reinforced some of the keys to performing effective audits.</i>
		Skill building (11)	<i>Yes, because I will be creating working papers of my own and potentially viewing others work.</i>
	No (3)	Too simple (1)	<i>I think the cases were good for future career use, but I feel that the comparative judgment was too simple.</i>
		No value in comparison (1)	<i>No, I don't see how comparing your paper to someone else's would be beneficial.</i>
		Not like real life (1)	<i>Not really, if I was comparing something at work a real person would make the judgment, not a software.</i>
	Maybe (6)	For training (1)	<i>Maybe starting off or for training [of] new employees.</i>

		Depends on career (2)	<i>Yes? I'm not sure I'll go into auditing, but I'm sure it'll be useful one day!</i>
		Unsure (2)	<i>Possibly, but not for sure.</i>
		Subject matter (1)	<i>Maybe, if there were actual answers to compare them too.</i>

Item 12: As you made judgments, did you feel you learned anything? What specifically?	Yes (18)	Learned aesthetics (9)	<i>Little adjustments to presentation make a difference.</i>
		Self-improvement (2)	<i>I did notice little things that others did on their cases that could have slightly improved the quality of my own cases.</i>
		Material/Subject (3)	<i>I became more familiar [with] the processes of auditing as I compared other students' work.</i>
		Correctness (1)	<i>yes, how to spot the papers that best met the requirements</i>
		Comparing (3)	<i>Yes, I learned how to compare and contrast different cases.</i>
	No (6)	Already knew (2)	<i>No, I didn't learn anything because I had already done the assignment and learned all that I could learn.</i>
		No comment [No evidence statement] (4)	<i>No.</i>
	Maybe (2)	Knew rank order (1)	<i>I think I would have learned more had I been able to see the results of how my paper did compared to others.</i>
		Unsure (1)	<i>Not really, I felt like I was just looking at a bunch of spreadsheets that I still did not fully understand. I guess I learned what looks better visually on the work papers.</i>

Item 13: What was the basis for your judgment decisions? What formed the criteria for your decision to choose one item over another?	Accuracy (6)	NA	<i>I mainly tried to look for correctness according to the instructor's rubric.</i>
	Appearance (4)	NA	<i>Mostly I just looked at formatting to see which one I liked better. I could have spent more time looking more in depth at the actual numbers and calculations.</i>
	Accuracy and Appearance (12)	NA	<i>I went off of the instructions that were provided. If both of the documents seemed to have the criteria correct then I would go off of the one that looked the most professional.</i>
	Self-Reference (2)	NA	<i>I compared the answers on other spreadsheets to my partner and I's spreadsheet.</i>

	Other Reference (1)	NA	<i>Use of information I had learned. Guidance from my professor.</i>
	None (1)	NA	<i>none</i>

#### 4. Discussion

Student performance on the specific assignments did not significantly improve throughout the course as a result of the ACJ intervention. However, there were several positive outcomes from implementing ACJ into the classroom experience. Students reported increased psychological safety in their class teams, particularly in their feelings that they could take risks and make mistakes in the presence of their team members (as reported in Item 1 of the survey) and that they would not be rejected because of their differences (as reported in Item 3 of the survey). However, an increase in psychological safety may or may not be a function of time. As evidenced by the survey results, students generally enjoyed the experience, and saw it as relevant to their future careers. Additionally, the instructor reported anecdotal evidence to support this point; while students attended a conference for recruiting mid-semester, visiting employers were excited that workpaper review practice was being added to the curriculum and this point was noted by both the students and the instructor. The ACJ intervention helped to facilitate engaging class discussion about preparing and reviewing audit workpapers as students were able to experience both aspects of the process. Based on student comments in the survey and in classroom discussion, students felt they were able to easily make connections to concepts taught in class and their application and importance to real-world practice.

There were a large number of student comments that pertained to the technical functioning and interface of the ACJ software. This study was an investigation into the process of peer feedback in accounting education and the influence this educational practice had on psychological safety and not into the technical mechanics used to convey that process. Therefore, a further treatment as to why there was a range of divergent views on the technical components of ACJ, is not within the scope of this paper. However, it can be noted that in this intervention, when there were limiting aspects or technical glitches to the process, students did report those issues over other elements of their learning experience.

Due to our methods in anonymous completion of the surveys, gender was one of the only demographic characteristics we received. Therefore, it became a default grouping variable for us to conduct our analysis. Even while the number of females outnumbered that of males, there was no statistically significant interaction for gender. Though, from the results, it may appear that males tended to score higher in psychological safety in the presurvey and females scored higher in the postsurvey, due to a lack of pairing, these findings need further investigation. During the presurvey, the answer “neutral” was used five times more than in the postsurvey. Though there is no statistical significance, we believe that there may be some practical significance contained in these findings. An example of such practicality may be that psychological safety is not easily diagnosed or attributed in a first impression but is developed over time and only over that period of time can there be certainty of exactly how

psychologically safe a team or team member feels. Further research is needed in this area to conclude this with any statistical significance. Above all, introducing an intervention that included peer review and feedback and a best-to-worst ranking of student work, did not negatively impact student psychological safety.

Our results with the ACJ intervention are surprising, as much of the ACJ literature has shown some significant advantage to those who use ACJ in formative assessment over those who do not. We made several observations that may shed light on the disparity in our findings from the extant literature.

#### *4.1 Dosage Effect*

There is some uncertainty in the literature surrounding ACJ as to what role the dosage effect may have on student performance. The question as to whether a student does better on a certain assignment simply because they have been exposed to it multiple times is unclear but is inherently part of the positive findings in ACJ research. In our study, we had the students split that dosage from beginning to middle of the semester and middle to end. One may expect to observe that the students that utilized ACJ from beginning-to-middle of the semester would outperform their peers only until the other class members received their ACJ dosage. Ultimately, then, by the end of the semester all students would be performing at the same level. This finding was also not present. However, what seems to be the case is that there may have been another dosage effect in place: as students became more familiar with the subject matter and the audit documentation process, the effect of ACJ was lessened, thus showing no statistically significant improvement among or between groups over time.

#### *4.2 Grading*

Although we initially expected that points awarded for the grades would be well aligned with the ACJ rank (i.e., those who ranked the highest in the ACJ sessions would also receive the highest grade), we discovered some unexpected results due to the method that was used in grading by the professor. Rather than grading the cases on the standard bell curve, which would have ranked students across a spectrum, the principle of concept mastery was the guiding factor in grading since that most closely matches the requirements in the professional standards for workpaper preparation in auditing. Therefore, all groups that met the required criteria in the rubric could receive full points for the case for purposes of their grade, even if they achieved a lower rank in the ACJ activity. For the case studies that were more simple, this created a scenario where most of the groups mastered the required concepts and were able to earn full or nearly full points, even if they were listed near the end on the ACJ rank. Therefore, the highest ACJ rank did not always correlate with highest points for grading, and vice versa for the lowest ACJ rank. We confirmed that basing grades on ACJ rank is more closely aligned with grading methods that use a standard bell curve rather than grading methods that use concept mastery as the objective.

#### *4.3 Professor Perspective*

Although ACJ rank does not necessarily correlate with academic grading for case studies in an auditing course due to the factors mentioned in the previous paragraph, ACJ does appear

to have real practical value in preparing students for careers in the auditing profession. As mentioned in the introduction, feedback from the profession indicates that one of the largest gaps between academic learning and preparation for required skills in the auditing profession is the aspect of workpaper preparation, documentation, and review. Additionally, student responses in survey items 10 and 11 indicated that the majority of students felt that this activity prepared them for the profession. Based on these student and professional responses, ACJ's primary value in auditing courses may not necessarily be found in improving academic grades, rather, it may lie in preparing students for future careers in the auditing profession by exposing them to the peer review and feedback processes required in the profession, and helping them begin to experience psychological safety and to develop the skills necessary to be successful in that area of daily professional work.

#### *4.4 Future Research*

As described above, ACJ has been implemented as an effective learning tool in settings ranging from elementary to university education. As students in an auditing course did not experience a greater improvement in grade performance relative to when traditional educational methods were used, we cannot conclude that ACJ is effective for significantly improving grades in the business education setting where spreadsheets are involved. However, before jumping to this conclusion we believe future research is warranted. Students reported enjoying the experience, and academic performance was in no way hindered by the ACJ experience. To our knowledge, this is the first attempt to use ACJ in business education with spreadsheets. With spreadsheets, in addition to elements of visual presentation, quantitative accuracy is also important. Future research could seek to disentangle these elements with respect to the effectiveness of ACJ as an educational tool. It could also be the case that the effectiveness of ACJ as a learning tool is a function of dosage or time. Perhaps students were not given sufficient experience using ACJ to translate to improved academic performance.

#### *4.5 Conclusion*

Audit documentation is vital to the auditing profession. To ensure documentation meets standards of sufficiency and appropriateness, the profession has employed the process of peer review and feedback. Audit team members may not only provide essential feedback relating to the documentation but also in regard to team member performance. Implementing peer review as an educational practice and learned outcome in audit education programs is therefore vital to preparing auditors who have the skills and experience to participate as an able-bodied peer reviewer.

As psychological safety becomes an increasingly researched phenomena (Edmondson, 2019), the role it plays in helping teams do their best work cannot be overstated. Auditors fresh out of higher education programs need to understand more than how to conduct an effective audit; they need to have had experience in developing and fostering psychological safety in teams. Failing to have cultivated the habit of speaking up or taking responsibility for mistakes, risks, or failures while also being able to be responsive and attentive to those who commit such errors or take such leaps of faith and fail, can be devastating to team and organizational performance. Thus, students must learn not only how to perform peer review, but also how to

maintain healthy intergroup dynamics that make the peer review truly productive.

This intervention sought to combine the innovative approach of Adaptive Comparative Judgment—a new and validated form of peer review and feedback—and psychological safety to learn more about the relationship of intentional peer review practices in an educational setting. Much of our results with ACJ are insignificant, showing that such an intervention may not significantly improve our student’s academic performance over others. However, the value of the intervention was not lost on the students; they saw the experience as a hands-on approach to learning skills that could help them better prepare for their future careers. Their perceptions were validated by recruiters and personnel from the field who were told of the intervention being done. This may be anecdotal, but often value is in the eye of the beholder. Students felt that such an intervention was important and were grateful to have engaged with an educational practice that they saw as beneficial and promotive of their career goals.

Psychological safety—particularly in areas of team acceptance while making mistakes and team acceptance for those seen as different—significantly increased during the time span of this intervention. In a westernized culture where comparison and ranking with a sort of didactic view (e.g., this student’s work is better than another) could be seen as a negative approach in a classroom setting no such negative impact was revealed in our findings. Psychological safety scores, particularly those in connection with being able to take risks that may lead to mistakes and accepting others regardless of differences, significantly improved over time. Thus, it can be said that, in the case of this study, ACJ peer review and feedback did not hinder students’ perceptions of being able to take interpersonal risks in their education and, may potentially help improve psychological safety. Thus, the ACJ process may enhance student experiences in learning auditing technique while also to some extent improving their capacity to provide and receive peer feedback through fostering and growing psychological safety. This should encourage instructors and researchers to continue to research and implement such interventions in their own disciplines.

## References

- Achterberg, M., van Duijvenvoorde, A. C., Bakermans-Kranenburg, M. J., & Crone, E. A. (2016). Control your anger! The neural basis of aggression regulation in response to negative social feedback. *Social cognitive and affective neuroscience*, *11*(5), 712-720. <https://doi.org/10.1093/scan/nsv154>
- Agoglia, C. P., Brazel, J. F., Hatfield, R. C., & Jackson, S. B. (2010). How Do Audit Workpaper Reviewers Cope with the Conflicting Pressures of Detecting Misstatements and Balancing Client Workloads? *Auditing: A Journal of Practice & Theory*, *29*(2), 27-43. <https://doi.org/10.2308/aud.2010.29.2.27>
- Andiola, L. M., Lambert, T. A., & Lynch, E. J. (2018). Sprandel, inc.: Electronic workpapers, audit documentation, and closing review notes in the audit of accounts receivable. *Issues in Accounting Education*, *33*(2), 43-55. <https://doi.org/10.2308/iace-52055>
- Asare, S. K., & McDaniel, L. S. (1996). The effects of familiarity with the preparer and task complexity on the effectiveness of the audit review process. *The Accounting Review*, *71*(2),

139-159. Retrieved August 25, 2020, from <http://www.jstor.org/stable/248443>.

Backof, A. G. (2015). The impact of audit evidence documentation on jurors' negligence verdicts and damage awards. *The Accounting Review*, 90(6), 2177-2204. <https://doi.org/10.2308/accr-51072>

Bamber, E. M., & Bylinski, J. H. (1987). The effects of the planning memorandum, time pressure and individual auditor characteristics on audit managers' review time judgments. *Contemporary Accounting Research*, 4(1), 127-143. <https://doi.org/10.1111/j.1911-3846.1987.tb00659.x>

Bamber, E. M., & Ramsay, R. J. (2000). The Effects of Specialization in Audit Workpaper Review on Review Efficiency and Reviewers' Confidence. *AUDTING: A Journal of Practice*, 19(2), 147-157. <https://doi.org/10.2308/aud.2000.19.2.147>

Baron, R. A. (1988). Negative effects of destructive criticism: Impact on conflict, self-efficacy, and task performance. *Journal of Applied Psychology*, 73(2), 199-207. <https://doi.org/10.1037/0021-9010.73.2.199>

Bartholomew, S. R. (2017). Assessing open-ended design problems. *Technology and Engineering Education Teacher*, 76(6), 13-17. Retrieved from <https://search.proquest.com/docview/1898631832?accountid=13360>

Bartholomew, S. R., & Jones, M. D. (2021). A systematized review of research with adaptive comparative judgement (ACJ) in higher education. *International Journal of Technology and Design Education*. <https://doi.org/10.1007/s10798-020-09642-6>

Bartholomew, S. R., & Yoshikawa, E. (2018). A systematic review of research around Adaptive Comparative Judgment (ACJ) in K-16 education. *2018 Council on Technology and Engineering Teacher Education: Research Monograph Series* [Monograph]. Retrieved from <https://doi.org/10.20161/ctete-rms.v1.c.1>

Bartholomew, S. R., Strimel, G. J., & Yoshikawa, E. (2019). Using adaptive comparative judgment for student formative feedback and learning during a middle school design project. *International Journal of Technology and Design Education*, 29(2), 363-385. <https://doi.org/10.1007/s10798-018-9442-7>

Bartholomew, S. R., Yoshikawa, E., & Strimel, G. (2018). Perceptions and reality: Analyzing student experiences in ranking self and peer work through Adaptive Comparative Judgment. In *PATT35 Athlone Institute of Technology, Athlone, Ireland 18-21 June, 2018*.

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. <https://doi.org/10.1080/14606925.2018.1560876>

Bartholomew, S. R., Mentzer, N., Jones, M., Farrington, S., Johnson, N., Mohandas, L., & Thorne, S. (2019). Learning by evaluating: Using adaptive comparative judgment. Nashville, Tennessee. *106th Mississippi Valley Technology Teacher Education Conference and the 57th Southeastern Technology Education Conference*.



- Beghetto, R. A. (2009). Correlates of intellectual risk taking in elementary school science. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 46(2), 210-223. <https://doi.org/10.1002/tea.20270>
- Bergmann, B., & Schaeppi, J. (2016). A data-driven approach to group creativity. *Harvard Business Review*. Retrieved from <https://hbr.org/2016/07/a-data-driven-approach-to-group-creativity>
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551-575. <https://doi.org/10.1080/0969594X.2018.1441807>
- Bransford, J. D., & Donovan, S. M. (2005). Scientific inquiry and how people learn. In S.M. Donovan & J.D. Bransford (Eds.), *How students learn: History, mathematics, and science in the classroom*, 397–420. Washington, DC: The National Academies Press.
- Brazel, J. F., Agoglia, C. P., & Hatfield, R. C. (2004). Electronic versus Face-to-Face Review: The Effects of Alternative Forms of Review on Auditors' Performance. *Accounting Review*, 79(4), 949-966.
- Brutus, S., & Donia, M. (2009, August). The multiple benefits of using a centralized peer evaluation system in a business school. In *Academy of Management Proceedings* (Vol. 2009, No. 1, pp. 1-6). Briarcliff Manor, NY 10510: Academy of Management. <https://doi.org/10.5465/ambpp.2009.44243019>
- Byrnes, J. P. (1998). *The nature and development of decision-making: A self-regulation model*. Hillsdale, NJ: Erlbaum.
- Byrnes, J. P., Miller, D., & Schafer, W. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, 125(3), 367-383. <https://doi.org/10.1037/0033-2909.125.3.367>
- Cajiao, J., & Burke, M. J. (2016). How instructional methods influence skill development in management education. *Academy of Management Learning & Education*, 15(3), 508-524. <https://doi.org/10.5465/amle.2013.0354>
- Canty, D. (2012). The impact of holistic assessment using adaptive comparative judgment of student learning (PhD Thesis). Retrieved from [https://ulir.ul.ie/bitstream/handle/10344/6766/Canty\\_2012\\_impact.pdf?sequence=6](https://ulir.ul.ie/bitstream/handle/10344/6766/Canty_2012_impact.pdf?sequence=6)
- Chambers, R., & McDonald, P. (2013). Cultivating soft skills: Nontechnical, qualitative attributes should be developed throughout every step of the audit recruitment, retention, and talent development life cycle. *Internal Auditor*, 70(3), 45-50. Retrieved from <https://go.gale.com/ps/anonymou?id=GALE%7CA335190083&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=00205745&p=AONE&sw=w>
- Clifford, M. M. (1991). Risk taking: Theoretical, empirical, and educational considerations. *Educational Psychologist*, 26(3-4), 263-297. <https://doi.org/10.1080/00461520.1991.9653135>
- Cubrich, M., Elicker, J. D., Sully De Luque, M., & Gabel-Shemueli, R. (2019, July). Peer Feedback-Seeking: Importance of Feedback Environment Referents and Individual Cultural

Values. In *Academy of Management Proceedings* (Vol. 2019, No. 1, p. 12273). Briarcliff Manor, NY 10510: Academy of Management. <https://doi.org/10.5465/AMBPP.2019.12273abstract>

Czaja, R. J., & Cummings, R. G. (2010). Competitive online case presentations. *The Accounting Educators' Journal*, 20(1).

De Stobbeleir, K., & Ashford, S. J. (2014). The Power of Peers: Antecedents and Outcomes of Peer Feedback Seeking Behavior. In *Academy of Management Proceedings* (Vol. 2014, No. 1, p. 14128). Briarcliff Manor, NY 10510: Academy of Management. <https://doi.org/10.5465/ambpp.2014.37>

DeNisi, A. S., Randolph, W. A., & Blencoe, A. G. (1983). Potential problems with peer ratings. *Academy of Management Journal*, 26(3), 457-464. Retrieved from [http://real-eod.mtak.hu/19/1/INFORM\\_07.pdf#page=167](http://real-eod.mtak.hu/19/1/INFORM_07.pdf#page=167)

Domingo, M., Petkova, A., & Lamm, E. (2016). Let's Be Frank: The Real Improvement in Student Teamwork Effectiveness Following Peer Feedback. In *Academy of Management Proceedings* (Vol. 2016, No. 1, p. 12754). Briarcliff Manor, NY 10510: Academy of Management. <https://doi.org/10.5465/ambpp.2016.90>

Donia, M., O'Neill, T. A., & Brutus, S. (2015). Peer feedback increases team member performance, confidence and work outcomes: A longitudinal study. In *Academy of Management Proceedings* (Vol. 2015, No. 1, p. 12560). Briarcliff Manor, NY 10510: Academy of Management. <https://doi.org/10.5465/ambpp.2015.21>

Drexler, J. A., Beehr, T. A., & Stetz, T. A. (2001). Peer appraisals: Differentiation of individual performance on group tasks. *Human Resource Management*, 40, 333-345. <https://doi.org/10.1002/hrm.1023>

Duhigg, C. (2016, February 25). What google learned from its quest to build the perfect team. The New York Times Magazine: Work Issue. Retrieved from [https://www.nytimes.com/2016/02/28/magazine/what-google-learned-from-its-quest-to-build-the-perfect-team.html?\\_r=0&referer=http://lnkd.in](https://www.nytimes.com/2016/02/28/magazine/what-google-learned-from-its-quest-to-build-the-perfect-team.html?_r=0&referer=http://lnkd.in)

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350-383. <https://doi.org/10.2307/2666999>

Edmondson, A. (2004). Psychological safety, trust, and learning in organizations: a group-level lens. In R.M. Kramer & K.S. Cook (Eds.), *Trust and distrust in organizations: Dilemmas and approaches* (pp. 239-272). New York: Russell Sage.

Edmondson, A. C. (2019). *The fearless organization: Creating psychological safety in the workplace for learning, innovation, and growth*. Hoboken, NJ: John Wiley & Sons.

Edmondson, A. C., Higgins, M., Singer, S., & Weiner, J. (2016). Understanding psychological safety in health care and education organizations: A comparative perspective. *Research in Human Development*, 13(1), 65-83. <https://doi.org/10.1080/15427609.2016.1141280>

Erez, A., Schilpzand, P., Leavitt, K., Woolum, A. H., & Judge, T. A. (2015). Inherently

relational: Interactions between peers' and individuals' personalities impact reward giving and appraisal of individual performance. *Academy of Management Journal*, 58(6), 1761-1784. <https://doi.org/10.5465/amj.2011.0214>

Fargher, N. L., Mayorga, D., & Trotman, K. T. (2005). A Field-Based Analysis of Audit Workpaper Review. *AUDITING: A Journal of Practice*, 24(2), 85-110. <https://doi.org/10.2308/aud.2005.24.2.85>

Fini, R., Jourdan, J., & Perkmann, M. (2018). Social valuation across multiple audiences: The interplay of ability and identity judgments. *Academy of Management Journal*, 61(6), 2230-2264. <https://doi.org/10.5465/amj.2016.0661>

Frank, M. L., & Hoffman, V. B. (2015). How Audit Reviewers Respond to an Audit Preparer's Affective Bias: The Ironic Rebound Effect. *Accounting Review*, 90(2), 559-577. doi:10.2308/accr-50873

Garvin, D. A., Edmondson, A. C., & Gino, F. (2008). Is yours a learning organization?. *Harvard business review*, 86(3), 109. Retrieved from <https://hbr.org/2008/03/is-yours-a-learning-organization>

Gielen, M., & de Wever, B. (2015). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior*, 52, 315-325. <https://doi.org/10.1016/j.chb.2015.06.019>

Greenberg, J., & Liu, C. C. (2015). Social relational foundations of peer evaluation. In *Academy of Management Proceedings* (Vol. 2015, No. 1, p. 11053). Briarcliff Manor, NY 10510: Academy of Management. <https://doi.org/10.5465/ambpp.2015.11053abstract>

Hamer, J., Purchase, H., Luxton-Reilly, A., & Denny, P. (2015). A comparison of peer and tutor feedback. *Assessment & Evaluation in Higher Education*, 40(1), 151-164. <https://doi.org/10.1080/02602938.2014.893418>

Harari, D., Lemoine, G. J., Parsons, C. K., & Sharp, J. S. (2015). The new"" How am I doing?"": Introduction and Development of a Team Feedback Climate Construct. In *Academy of Management Proceedings* (Vol. 2015, No. 1, p. 17398). Briarcliff Manor, NY 10510: Academy of Management. <https://doi.org/10.5465/ambpp.2015.17398abstract>

Harding, N. (2010). Understanding the structure of audit workpaper error knowledge and its relationship with workpaper review performance. *Accounting & Finance*, 50(3), 663-683. <https://doi.org/10.1111/j.1467-629X.2010.00342.x>

Higgins, M., Ishimaru, A., Holcombe, R., & Fowler, A. (2012). Examining organizational learning in schools: The role of psychological safety, experimentation, and leadership that reinforces learning. *Journal of Educational Change*, 13(1), 67-94. <https://doi.org/10.1007/s10833-011-9167-9>

Howorth, C., Smith, S. M., & Parkinson, C. (2012). Social learning and social entrepreneurship education. *Academy of Management Learning & Education*, 11(3), 371-389. <https://doi.org/10.5465/amle.2011.0022>

Jurado, J. A. (2011). Group projects in interior design studio classes: Peer feedback benefits.

*Journal of Family and Consumer Sciences*, 103(1), 34. Retrieved from <https://search.proquest.com/docview/887107513?accountid=13360>

Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal*, 33(4), 692-724. <https://doi.org/10.5465/256287>

Kessel, M., Kratzer, J., & Schultz, C. (2012). Psychological safety, knowledge sharing, and creative performance in healthcare teams. *Creativity and Innovation Management*, 21(2), 147-157. <https://doi.org/10.1111/j.1467-8691.2012.00635.x>

Kidwell, R. E., & Bennett, N. (1993). Employee propensity to withhold effort: A conceptual

Kim, Y. J., & Kim, J. (2020). Does negative feedback benefit (or harm) recipient creativity? The role of the direction of feedback flow. *Academy of Management Journal*, 63(2), 584-612. <https://doi.org/10.5465/amj.2016.1196>

Kimbell, R. (2007). E-assessment in project e-scape. *Design & Technology Education: An International Journal*, 12(2), 66-76. Retrieved from [https://www.ariadne.ac.uk/DATE/article/download/Journal\\_12.2\\_0707\\_RES6/59](https://www.ariadne.ac.uk/DATE/article/download/Journal_12.2_0707_RES6/59)

Kimbell, R. (2012). Evolving project e-scape for national assessment. *International Journal of Technology & Design Education*, 22(2), 135-155. <https://doi.org/10.1007/s10798-011-9190-4>

Kimbell, R. (2018). Constructs of quality and the power of holism. In *PATT35 Athlone Institute of Technology, Athlone, Ireland 18-21 June, 2018*. Retrieved from [shorturl.at/pALRW](http://shorturl.at/pALRW)

Lambert, T. A., & Agoglia, C. P. (2011). Closing the loop: Review process factors affecting audit staff follow-through. *Journal of Accounting Research*, 49(5), 1275-1306. <https://doi.org/10.1111/j.1475-679X.2011.00423.x>

Li, L., & Gao, F. (2016). The effect of peer assessment on project performance of students at different learning levels. *Assessment & Evaluation in Higher Education*, 41(6), 885-900. <https://doi.org/10.1080/02602938.2015.1048185>

Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology*, 41(3), 525-536. <https://doi.org/10.1111/j.1467-8535.2009.00968.x>

Mayo, M., Kakarika, M., Pastor, J. C., & Brutus, S. (2012). Aligning or inflating your leadership self-image? A longitudinal study of responses to peer feedback in MBA teams. *Academy of Management Learning & Education*, 11(4), 631-652. <https://doi.org/10.5465/amle.2010.0069>

McCarthy, J. (2017). Enhancing feedback in higher education: Students' attitudes towards online and in-class formative assessment feedback models. *Active Learning in Higher Education*, 18(2), 127-141. <https://doi.org/10.1177/1469787417707615>

Metzgar, M. (2016). Using adaptive comparative judgement to assess student work in an MBA course. *International Journal for Infonomics*, 9(3), 1217-1219.

<https://doi.org/10.20533/iji.1742.4712.2016.0148>

model to intersect three avenues of research. *Academy of Management Review*, 18, 429-456.

<https://doi.org/10.5465/amr.1993.9309035146>

Morales, S., Vallorani, A., & Pérez-Edgar, K. (2019). Young children's behavioral and neural responses to peer feedback relate to internalizing problems. *Developmental cognitive neuroscience*, 36, 100610. <https://doi.org/10.1016/j.dcn.2018.12.008>

Morrison, E., Hutcheson, S., Nilsen, E., Fadden, J., & Franklin, N. (2019). *Strategic doing: ten skills for agile leadership*. Hoboken, NJ: John Wiley & Sons.

Newman, A., Donohue, R., & Eva, N. (2017). Psychological safety: A systematic review of the literature. *Human Resource Management Review*, 27(3), 521-535. <https://doi.org/10.1016/j.hrmr.2017.01.001>

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education*, 39(1), 102-122. <https://doi.org/10.1080/02602938.2013.795518>

Oakley, B., Felder, R. M., Brent, R., & Elhajj, I. (2004). Turning student groups into effective teams. *Journal of Student-Centered Learning*, 2(1), 9-34. Retrieved from [shorturl.at/abDEI](http://shorturl.at/abDEI)

Ohland, M. W., Loughry, M. L., Woehr, D. J., Bullard, L. G., Felder, R. M., Finelli, C. J., ... Schmucker, D. G. (2012). The comprehensive assessment of team member effectiveness: Development of a behaviorally anchored rating scale for self-and peer evaluation. *Academy of Management Learning & Education*, 11(4), 609-630. <https://doi.org/10.5465/amle.2010.0177>

Payne, E. A., & Ramsay, R. J. (2008). Audit documentation methods: A path model of cognitive processing, memory, and performance. *AUDTING: A Journal of Practice*, 27(1), 151-168. <https://doi.org/10.2308/aud.2008.27.1.151>

Payne, E. A., Ramsay, R. J., & Bamber, E. M. (2010). The Effect of Alternative Types of Review on Auditors' Procedures and Performance. *AUDTING: A Journal of Practice*, 29(1), 207-220. <https://doi.org/10.2308/aud.2010.29.1.207>

Piper, A. (2017). The dynamics of interpersonal behavior: To be successful, auditors need to cultivate their soft skills just as much as their technical abilities. *Internal Auditor*, 74(3), 50-56.

Retrieved

from

<https://go.gale.com/ps/anonymous?id=GALE%7CA497672680&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=00205745&p=AONE&sw=w>

Plant, K., Barac, K., & Sarens, G. (2019). Preparing work-ready graduates—skills development lessons learnt from internal audit practice. *Journal of Accounting Education*, 48, 33-47. <https://doi.org/10.1016/j.jaccedu.2019.06.001>

Pollitt, A. (2015). *On 'Reliability' bias in ACJ*. Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment. <https://doi.org/10.13140/RG.2.1.4207.3047>

Pollitt, A. (Producer). (2004). Let's stop marking exams. Retrieved from

<http://www.cambridgeassessment.org.uk/images/109719-let-s-stop-marking-exams.pdf>

Potter, T., Englund, L., Charbonneau, J., MacLean, M. T., Newell, J., & Roll, I. (2017). ComPAIR: A new online tool using adaptive comparative judgement to support learning with peer feedback. *Teaching & Learning Inquiry*, 5(2), 89-113. <https://doi.org/10.20343/teachlearninqu.5.2.8>

Public Company Accounting Oversight Board (PCAOB). (2016, December 31). Auditing Standards (AS 1215.02). Retrieved from <https://pcaobus.org/Standards/Auditing/Pages/AS1215.aspx>

Public Company Accounting Oversight Board (PCAOB). (2016, December 31). Auditing Standards (AS 1201.05). Retrieved from <https://pcaobus.org/Standards/Auditing/Pages/AS1201.aspx>

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).

Rosman, A., Biggs, S., Graham, L., & Bible, L. (2007). Successful Audit Workpaper Review Strategies in Electronic Environments. *Journal of Accounting, Auditing & Finance*, 22(1), 57-83. <https://doi.org/10.1177/0148558X0702200105>

Rozovsky, J. (2015). The five keys to a successful Google team. Retrieved from <https://rework.withgoogle.com/blog/five-keys-to-a-successful-google-team/>

Rynes, S. L., Trank, C. Q., Lawson, A. M., & Ilies, R. (2003). Behavioral Coursework in Business Education: Growing Evidence of a Legitimacy Crisis. *Academy of Management Learning & Education*, 2(3), 269-283. <https://doi.org/10.5465/amle.2003.10932135>

Schein, E. H., & Bennis, W. (1965). *Personal and organizational change through group methods*. New York: Wiley.

Schepers, J., de Jong, A., Wetzels, M., & de Ruyter, K. (2008). Psychological safety and social support in groupware adoption: A multi-level assessment in education. *Computers & Education*, 51(2), 757-775. <https://doi.org/10.1016/j.compedu.2007.08.001>

Seery, N., & Canty, D. (2017) Assessment and learning: The proximal and distal effects of comparative judgment. In: de Vries M. (Editor) *Handbook of Technology Education*. Springer International Handbooks of Education. Springer, Cham. [https://doi.org/10.1007/978-3-319-38889-2\\_54-1](https://doi.org/10.1007/978-3-319-38889-2_54-1)

Seery, N., Buckley, J., Delahunty, T., & Canty, D. (2019). Integrating learners into the assessment process using adaptive comparative judgement with an ipsative approach to identifying competence based gains relative to student ability levels. *International Journal of Technology and Design Education*, 29(4), 701-715. <https://doi.org/10.1007/s10798-018-9468-x>

Shankar, P. G., & Tan, H. T. (2006). Determinants of audit preparers' workpaper justifications. *The Accounting Review*, 81(2), 473-495.

- Sluijsmans, D., Dochy, F., & Moerkerke, G. (1998). Creating a learning environment by using self-, peer- and co-assessment. *Learning environments research*, 1(3), 293-319. <https://doi.org/10.2308/accr.2006.81.2.473>
- Stefani, L. A. J. (1994). Peer, self and tutor assessment: relative reliabilities. *Studies in Higher Education*, 19(1), 69-75. <https://doi.org/10.1080/03075079412331382153>
- Stewart, G. L., Courtright, S. H., & Barrick, M. R. (2009, August). Peer-based reward and individual performance: a field examination. In *Academy of Management Proceedings* (Vol. 2009, No. 1, pp. 1-7). Briarcliff Manor, NY 10510: Academy of Management. <https://doi.org/10.5465/ambpp.2009.44247137>
- Sweeney, J. T., Suh, I. S., Dalton, K. C., & Meljem, S. (2017). Are workpaper reviews preparer-specific?. *The British Accounting Review*, 49(6), 560-577. <https://doi.org/10.1016/j.bar.2017.09.003>
- Tan, H. T., & Trotman, K. T. (2003). Reviewers' responses to anticipated stylization attempts by preparers of audit workpapers. *The Accounting Review*, 78(2), 581-604. <https://doi.org/10.2308/accr.2003.78.2.581>
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34(4), 273-286. <https://doi.org/10.1037/h0070288>
- Ulrich, T. A., & Blouch, W. E. (2018). Cpa Practitioners' Feedback on Auditing Curriculum. *Journal of Business & Accounting*, 11(1), 68-96. Retrieved from [http://asbbs.org/files/2019/JBA\\_Vol\\_11.1\\_Fall\\_2018.pdf#page=68](http://asbbs.org/files/2019/JBA_Vol_11.1_Fall_2018.pdf#page=68)
- Zhang, L. (2019). Investigating Differences in Formative Critiquing between Instructors and Students in Graphic Design (Master's thesis). Retrieved from Purdue University Graduate School. <https://doi.org/10.1080/14606925.2018.1560876>
- Zhang, Y., Fang, Y., Wei, K. K., & Chen, H. (2010). Exploring the role of psychological safety in promoting the intention to continue sharing knowledge in virtual communities. *International Journal of Information Management*, 30(5), 425-436. <https://doi.org/10.1016/j.ijinfomgt.2010.02.003>

### **Copyright Disclaimer**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).