# Investigating Cognitive Discrepancies Between Verbalized and Written Feedback in 9th-Grade Engineering Students' Evaluation Processes


## Abstract

Research using the Thinking Aloud (TA) process has demonstrated its effectiveness in revealing students' cognitive strategies and providing valuable insights into their thought processes during problem-solving and evaluation tasks. Similarly, Learning by Evaluating (LbE) has been shown to enhance students' critical thinking and judgment skills as they compare and assess artifacts. However, discrepancies are believed to exist between what students verbalize during TA and what they articulate in written feedback. Our study explores this difference in thinking and writing among 9th-grade engineering students while evaluating engineering design artifacts. Investigating potential differences in thinking and writing is essential, especially if we hope to use writing to gauge their cognitive abilities, such as evaluative thinking and decision-making.

This study employs qualitative methods utilizing TA and LbE as data collection settings. Ninth-grade students from three schools in Dekalb County, GA were asked to think aloud while comparing artifacts, sharing their reasoning and the factors they consider when making a judgment. Verbalized thoughts and then written responses were coded according to a predetermined codebook, then compared for whether codes exist in both responses. Initial findings noted some refinement in reasoning within the written responses, but generally few points of evidence when reasoning. These findings emphasize the need to enhance students' critical thinking skills and improve their ability and opportunities to express reasoning and judgments. Future research may focus on identifying specific cognitive barriers and developing interventions that support the alignment of students' verbal and written evaluation processes.



***Keywords***: critical thinking, Learning by Evaluating, thinking aloud, cognition



Presenters: Daniel Bayah (dkb85825@uga.edu), Andrew Jackson (andrewjackson@uga.edu), Nathan Mentzer (nmentzer@purdue.edu), Scott Bartholomew (scottbartholomew@byu.edu)

## Introduction

The first author, Daniel, reflected:

> *Under my father's guidance as a teacher, I vividly recall the significant impact of thinking aloud (TA) during elementary school activities such as reading and mathematics. This practice notably improved my comprehension and retention of academic material throughout my educational journey. Verbalizing thoughts in these early stages enhanced my cognitive processes and nurtured critical thinking skills—an influence that continues to shape my understanding of teaching and learning today. This background serves as a foundation in the current context of evaluating 9th-grade engineering students. My experience with TA reinforces the importance of verbalizing thought processes as a powerful cognitive tool. Just as TA was instrumental in my learning process, and might facilitate students' thinking, it also holds the potential to reveal the depth of their thought processes: for teachers or researchers, the TA process provides insights into how students reason and judge that we might not otherwise be able to attain.*

Given that the basis of much engineering design assessment is on students' written documentation (Kelley, 2011), understanding differences between students' thinking abilities and what ends up in their written materials is crucial for developing effective teaching strategies. In technology and engineering education, students are often required to evaluate complex design projects (Goel & Pirolli, 1992). Making decisions based on evidence is a key feature of successful design (Crismond & Adams, 2012), as is argumentation, which involves formulating and defending claims (Wilson-Lopez et al., 2020). The alignment (or misalignment) between thinking and written feedback can affect how well students proceed in design. For example, students' abilities to articulate their reasoning, apply critical thinking, and communicate their judgments effectively, play a role in the effectiveness of their final solutions (Atman et al., 1999; Mentzer et al., 2015; Sampson & Clark, 2008).

Previous research indicates the significance of this challenge in educational evaluation, in general. Ericsson and Simon (1993) showed that verbalization can reveal complex reasoning that may be difficult for students to condense into written responses due to limitations like cognitive overload or memory constraints (Sweller, 1988). Furthermore, Sadler (1989) noted that peer assessment is only effective when students' feedback accurately reflects their thinking. However, the accuracy of these comments are compromised by differences in verbal and written forms of feedback. Written feedback often appears less detailed or reflective of students' authentic thought process. Understanding the reasons behind these differences—whether due to cognitive overload (Sweller, 1988), lack of vocabulary, or other factors—can help educators develop targeted strategies to bridge the gap between students' thinking and written expressions.

In this study, we investigated how 9th-grade students transition from thinking to written evaluation forms. We were guided by the research question, "What cognitive discrepancies emerge between thinking and written feedback in 9th-grade engineering students?" We conducted the study within a Learning by Evaluating (LbE) context, described next. Our own observations of student thinking and written reasoning in the LbE context initially motivated the

study. This rich setting will help uncover potential cognitive barriers or filters that cause variations between what students think and say and write, offering insights into their design thinking. This more nuanced understanding of students' design thinking may also be used to improve teaching practices and support students in both forms of reasoning.

## Background

### Learning by Evaluating (LbE)

LbE is a novel educational approach that has been applied to design thinking (Bartholomew et al., 2023; Mentzer et al., 2023). In the context of LbE, students engage with artifacts or examples by comparing, evaluating, and critically reflecting on their decisions. Once the pedagogical strategy is introduced, students are presented with several artifacts to compare and contrast, making an evaluation about which is better according to a teacher-specified criteria called the *holistic statement*. LbE emphasizes both thinking and writing, in that students' incipient arguments are mentally reasoned through comparison of the artifacts and then documented in a *judgment statement.* Following students' individual comparisons, class-wide discussion allows for elicitation of students' judgment, identification of patterns of quality among the design artifacts seen, and class-wide consensus about indicators of quality for the given design context.

We have previously argued that LbE experiences leverage several epistemic practices used to support engineering argumentation (Jackson et al., 2022; Jackson et al., 2023; Wilson-Lopez et al., 2020). In particular, using "differing texts" is a noted literacy practice—we interpret this to include showing different artifacts for analysis and critique—and "whole-class discussion" is among the oral language practices recommended to help students construct their perspectives and arguments.

LbE also shifts two fundamental aspects of traditional design pedagogy. First, instead of relying on teacher-centered evaluation, students critically assess example materials. By engaging students as evaluators, the approach empowers students to make judgments based on design quality criteria and fosters active learning (Chi & Wylie, 2014). The developmental stage of 9th-grade students makes LbE a fitting strategy. At this age students are transitioning from concrete to more abstract thinking. LbE may help bridge this transition by requiring students to justify their decisions, moving them from surface-level observation to deeper, more conceptual reasoning.

Second, unlike summative or formative evaluation, LbE introduces evaluation at the beginning of the learning process. By comparing existing examples early on, students can construct a deeper understanding of the project's goals and essential elements before they initiate their work (Sadler, 1989). This approach amplifies part of the intent of formative assessment practices that promote self-regulated learning (Nicol & Macfarlane-Dick, 2006). The need to discriminate between artifacts with different, sometimes conflicting, strengths encourages students to navigate competing values, deepening their understanding of design as a multi-faceted and subjective process (Cross, 2006). Once students select their preferred alternative, they must justify their decision through written explanations grounded in evidence

and reasoning, which further elicits critical thinking (Ennis, 1985). This practice strengthens their ability to articulate their design judgments and connects to the broader goals of design education, emphasizing the importance of reflective practice and critical evaluation (Kimbell, 2012). Other research also provides evidence that examples and comparison are helpful in student learning (Sadler, 1989; van Hattum-Janssen & Lourenço, 2006).

Although we have hoped to understand students' reasoning process through their written comments, in person observations have made us wonder whether these written comments satisfactorily capture students' level of critical thinking. Analysis of students' comments has highlighted their brevity, and that they frequently report a single justification for making the decision (Jackson et al., 2023; Mentzer et al., 2023; Mentzer et al., 2024). This stands in contrast to the quality of discussions that we have observed and that participating teachers report (Lee et al., 2024). The investigation herein is to gather empirical evidence about students' evaluative thinking and written comments.

## Differences Between Thinking and Written Feedback

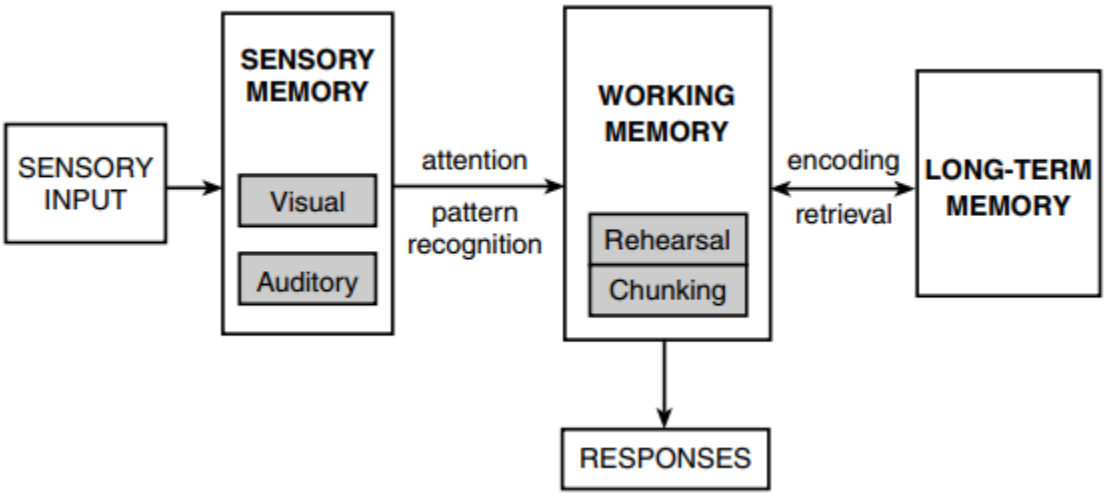
Although there are intuitive reasons to expect a difference between thinking and written feedback, empirical evidence supports this argument as well (Pugalee, 2004). According to Sweller (1988), writing acts as a "discovery process," helping students consolidate and refine their thinking. However, cognitive barriers or filters may come into play, leading to differences between thinking and writing. Several factors, including cognitive load, memory recall, and task complexity can affect the transition between what students think and what they write (Sweller, 1988). We identified several theoretical views which frame the transition between thinking and writing and support expectations of a difference in students' evaluative output.

### *Cognitive Information Processing Theory*

A first approach that infers a difference between students' thinking and writing is Cognitive Information Processing Theory (CIPT). According to CIPT, learning and decision-making involve several stages including attention, encoding, storage, and retrieval (Atkinson & Shiffrin, 1968). The production of a given response involves these stages and assumes that the learner has obtained and stored adequate information through these stages, and then is able to retrieve it. A visual overview of the relationships in CIPT from Driscoll (2005) is given in Figure 1. The effects of various elements of cognitive information processing theory have been explored in detail, related to writing.

**Figure 1**

*Cognitive Information Processing Theory Diagram*




It is recommended that learning activities consider the limits of working memory and aim to reduce unnecessary mental strain, called cognitive load (Kellogg, 2001; Sweller, 1988). Kellogg (2001) examined how working memory affects writing processes, focusing on planning, sentence generation, and reviewing steps. They suggested that working memory is essential for skilled writing because it enables writers to manage several tasks simultaneously, such as organizing ideas while drafting. Swanson and Berninger (1996) explored how specific components of working memory impacted writing tasks. Their research supports the view that limitations in working memory restrict writing performance, suggesting the importance of cognitive interventions to support students' capacities. Similarly, writers with greater capacity for cognitive load are able to produce more coherent and organized text. Considering cognitive load has influenced strategies to make complex information more accessible for learners.

Flower and Hayes (1981) reiterated that the cognitive processes involved in writing support the idea that writing is not just about transcribing thoughts but involves significant planning, decision-making, and iterative processes. Subsequently, written judgments after verbalized thoughts involve a more complex cognitive operation—one that does not completely represent the reflections made during the thinking phase. Hayes et al. (1987) considered the work of revising writing through CIPT, and argued the centrality of information retrieval from memory. Once the writer activates prior knowledge, and produces writing, the writing can be evaluated in the context of the task, with recursive returns to the act of generating writing (Berninger et al., 2009). In this way, there is a back-and-forth between pre-writing, drafting, and revising (with information moving from attention and long-term memory to working memory), similar to our view of the design thinking process.

Looking at the LbE experience through this lens, the attention or focus on specific aspects of the design examples in LbE may connect with long-term memory (what students have

learned from their classes or personal experiences) about what matters in design. These connections are then encoded as a written response, capturing their analysis and insights. The thinking involves more immediate cognitive engagement, reflecting on and processing information in real time. On the other hand, written feedback requires students to retrieve, organize, and articulate information after this cognitive processing. A potential discrepancy may arise because of the additional steps between thinking and writing (e.g., retrieval from memory, language processing, and articulation; Baddeley, 2000).

### *Models on Learning to Write*

In studying literacy and writing, other scholars have taken a more conceptual model (in contrast to the cognitive model) to describe how writing develops. Several evolutions of these models describe an increasingly complex view on what leads up to the writing students ultimately produce.

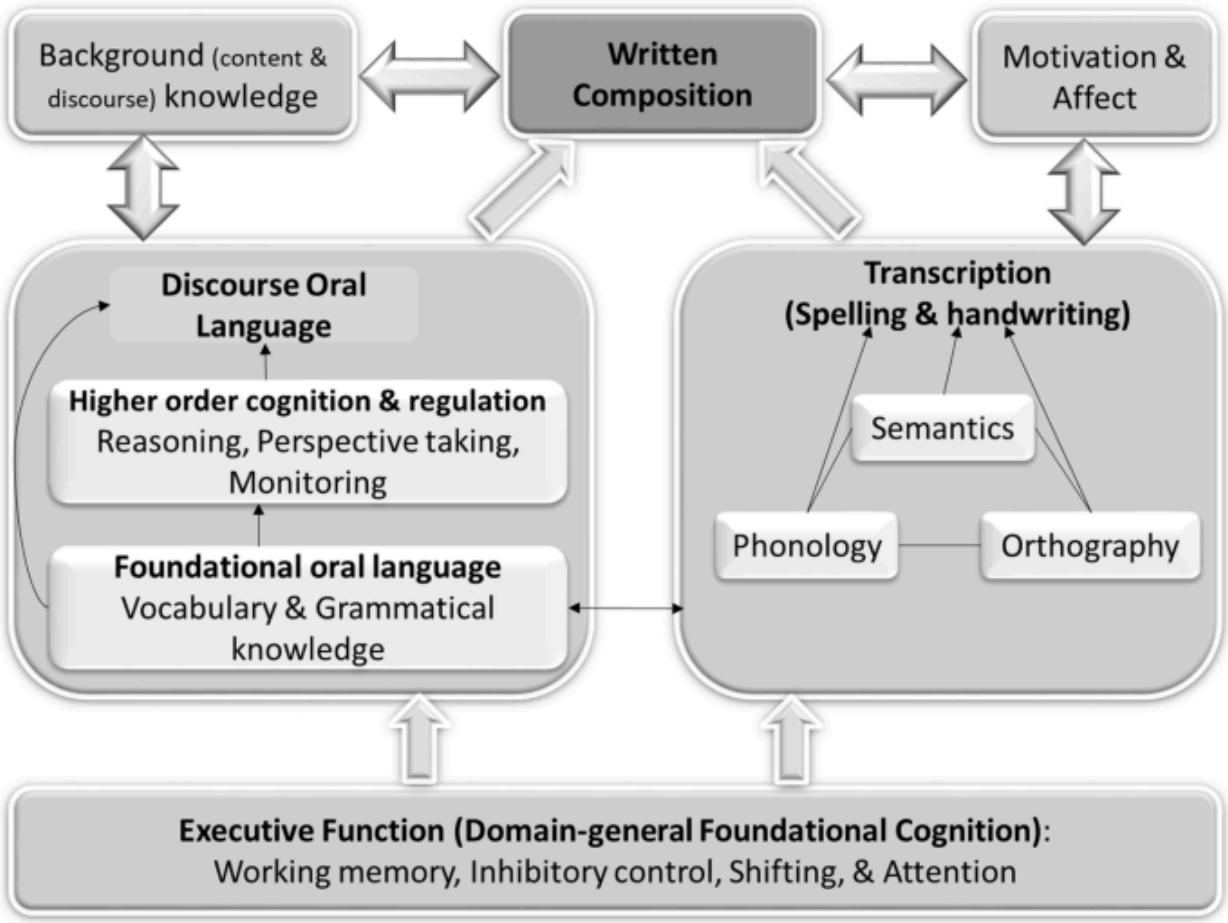
The Simple View of Writing (Berninger et al., 2002) posits that writing primarily comprises two key components: transcription (i.e., spelling, handwriting, or typing) and ideation (i.e., expressing thoughts and ideas). Because younger writers struggle with transcription, for example, this limits their ability to focus on higher-order writing processes like planning and revising. The Not-So-Simple View of Writing (Hayes & Flower, 1980) expanded on this model to acknowledge the complexity of writing, specifically including phases such as planning, revising, and self-regulation. This model emphasizes the recursive nature of writing (drawing on aspects of self-regulation) and recognizes that skilled writers often return to earlier phases of writing to improve their work. This dynamism in writing could be influenced by better understanding of the task or audience, or in response to previously written elements. Other research has also demonstrated the role of oral language in supporting writing (Kim & Park, 2019).

Further models such as the Direct and Indirect Effects Model of Writing (DIEW) further elaborated the factors believed to influence writing development (see Figure 2). This model describes cognitive skills in addition to those in the Not-So-Simple View of Writing (e.g., perspective taking, monitoring), the writers' background knowledge and language skills, and their motivation and affect (Kim & Graham, 2021). This model recognizes that writing builds on multiple layers of cognitive and linguistic development, yet is a separate, complex activity from thinking or oral language. Taken together, these conceptual models organize writing as "an extremely complex skill" (Graham, 2019, p. 281), distinct from thinking.

If we look at the LbE experience from these views, the details that students produce in their written comments are based on layers of complexity. In LbE, their written comments are influenced by their current reflections and draw on oral language skills, previous knowledge, and critical insights shaped by their comparisons, deepening their understanding and engagement with the material. While the ways that writing emerges from the LbE experience, in combination with these theoretical views, supports the expectation of a difference between thinking and writing—not everything students think will be written—the reality and magnitude of this difference is unknown.

**Figure 2**

*The Direct and Indirect Effects Model of Writing (DIEW)*




## Methods

To explore the nature of students' thinking and writing, a qualitative design was employed. We leverage the LbE context and TA method to investigate potential cognitive discrepancies between thinking and written feedback during an engineering evaluation task. The qualitative design was chosen because it enables a rich understanding of the students' thought processes, and in-depth exploration of their critical thinking and decision making (Creswell & Poth, 2018). Participants in the study were 9th-grade students from three public high schools in DeKalb County, GA. The sample was purposefully selected from among those in a larger study (Jackson et al., 2023) to capture thinking and writing for closer analysis. Four students who had assented to the study and received parental consent to participate were recommended by teachers for their likely comfort with thinking out loud. Such comfort to participate and likelihood to share insight while thinking aloud are common selection criteria in TA studies (Ericsson & Simon, 1993).

## Data Collection

Data collected for the study was based on two complementary sources provided by students during the LbE experience, as previously described. First, students were asked to think aloud while evaluating the artifacts. Second, their written comments from the LbE session were gathered. Tashakkori and Teddlie (2003) emphasized that combining verbal and written data can provide deeper insights into cognitive processes. The types of written comments from LbE (or adaptive comparative judgment, which is the underlying process) used here are nearly always used to gain insight into the comparison process (e.g., Bartholomew et al., 2019; Jackson et al., 2022). However, to our knowledge, the combination of LbE and TA data is unique to this study.

### *Thinking Aloud (TA) Data*

The purpose of the TA protocol was to elicit students' thinking. TA is widely used in design and writing research (e.g., Atman et al., 2007; Ericsson & Simon, 1993). The method allows researchers to capture cognitive strategies in a naturalistic setting. Prior studies have highlighted the protocol's effectiveness in understanding decision-making in educational contexts (Chi, 1997), making it ideal for this research. Through the LbE session, students were presented with several pairs of engineering design artifacts and asked to verbally articulate their thought processes and decision making as they compared the items. Because the LbE experience is partly computer-based, students explained details of the artifacts they saw on the screen, and described a rationale for choosing between them. Students were set up with an individual microphone, which was later downloaded and transcribed. During the experience, a researcher also observed the student—this allowed us to encourage students to keep speaking if they were quiet and to take observation notes about the students' practices when making the evaluations.

### *Learning by Evaluating (LbE) Data*

After the initial thinking about the design artifacts, students were required to provide written feedback on each decision. The students recorded their comments in the RM Compare interface, which was used for the experience. Then we extracted and linked the session reports, including all the written feedback that students provided while evaluating the design artifacts. Once the reports were downloaded, we reconstructed the sequence of comparisons that each student was tasked with. This allowed us to revisit the context of each comparison, and understand verbal and written references to the artifacts. Using the session reports we also isolated the comments from each student who participated in the session we observed.

## Analysis

Analysis and interpretation began informally from the moment of observation, through the ways we completed observer notes and began to make sense of the experience in discussion. Yet, once the transcripts and student comments were obtained, analysis formally began by coding data according to a predetermined codebook that we have used to analyze LbE comments in prior research studies (Bartholomew et al., 2019; Jackson et al., 2022). Codes relate to key design criteria and rationale in engineering decision making, such as aesthetics, emphasizing the design process, or highlighting usability in the solution. (For a full description of the codes

applied and their original sources, see Jackson et al., 2022.) Coding of data was completed by two members of the research team individually, who then met to negotiate to consensus. Each form of student evaluation data was coded separately to distinguish between the thinking and writing that occurred.

Following coding, data for each student was reviewed holistically in discussion with the research team, with the codes treated as part of the story. We considered the contents of the transcript, written comments, and the labels we had applied throughout the coding process. This review paid special attention to the trajectory of students' work, from thinking to writing. For example, we examined the duration and quality of students' thinking—based on the transcript—and the focus of their thinking—based on the codes applied. Then, we examined the relative nature of their written comments—what was the quality of their comments and were the same topics mentioned? Based on these assessments we relay the nature of each students' experience when making comparisons and illustrate several patterns seen when looking across these cases.

## Results

### Student profiles

#### *Paul*

Paul was confident in his abilities, a fact evidenced by his willingness to participate in TA and share his justifications for each decision. The comparisons he worked on were from an LbE session prepared specifically for an upcoming project focused on student-led presentations. Paul's teacher had gathered examples from the past several years as students completed the project. It was relayed to the class that the artifacts were from previous years' students and that they related directly to the upcoming project, raising the relevance of making the comparisons. Having been gathered over time, there were numerous artifacts in the collection that might have been presented to each student. The session was also set up by the teacher to require many comparisons, suggesting that by the end of the comparisons students would see most images.

In combination to the large number of comparisons that students were expected to make, however, one other setting in the LbE session management led to difficulty in understanding potential differences in Paul's thinking and writing: only a few written comparisons were required to be made by students in the judgment process (and the teacher mentioned this). Accordingly, after making written comments on the first several comparisons, Paul's later arguments were only explained verbally as he thought out loud. While we did have some overlap in observing this thinking and writing, our ability to understand potential gaps in thinking and written arguments is obviously limited here. Said another way though, a clear takeaway might be that to enhance students' written comments, in general circumstances they must be made mandatory. This directly aligns with the self-regulatory or motivational aspects of the writing development models mentioned before.

When characterizing Paul's approach overall, we did observe a decline in the quality of comments, both over time (even in the written comments before they were no longer required) and relative to the spoken rationales. Although several different topics were mentioned out loud

when comparing earlier artifacts, these were not captured in Paul's written rationale. Furthermore, as Paul went through the numerous comparisons, the pace at which he made them accelerated. There were periods of his thinking where the only utterance was identifying the artifacts and then stating the judgment without any reasoning. On the one hand, it is possible that his first steps in identifying the artifact was sufficient to activate it in his working memory and he was able to quickly scrutinize differences in quality. In our judgment, this claim is also dubious given the surface-level characteristics that Paul tended to focus on in these later comparisons. We feel that more likely explanations for this pattern relate to fatigue in the process, having maximized cognitive load, or expectations about how to make the comparisons. However, the lack of verbal justification given is not enough to fully understand Paul's rationale for each decision.

***Amina***

Amina worked on a session related to brainstorming documentation and quality. Artifacts in this session had also been gathered from prior student work, meaning that she was familiar with the expectations and contexts from which they had been curated. Only three comparisons were required to be made by each student. Amina was thorough in her work and dealt with the comparisons with a deeper understanding of what the brainstorming process entailed. In fact, she took the longest per comparison of any of the students we observed, ensuring that each artifact was carefully examined before deciding between them.

Her thinking articulated a range of different justifications, including relatively deeper justifications such as discernment of the problem definition in the underlying work. Many of these coded items from thinking remained in her written evaluation as well. To illustrate the range of coded verbal comments, quality of her written comments, and the relative similarity between these forms of evaluation an example is provided in Table 1. On the left, several key phrases are excerpted from the thinking that occurred while making the comparison. They are highlighted according to the various codes that we agreed were present. On the right, the written comment submitted with the judgment is given and also annotated.

**Table 1**

*Example coding of of Amina's spoken and written comparison*

| Excerpted Thinking | Written Judgment |
|---|---|
| A would be considered better brainstorming design [Code: Brainstorming] as it defines the problem [Code: Problem Id] and explains what specifically they're doing to achieve the problem. Meanwhile, B [Code: Evaluation] is simply an image of what they wanted to do...<br>So I chose A because A... actually explained what they're trying to do to solve it [Code: Communication], ... which is what we were supposed to do [Code: Completeness]. | A is better, as it actually defines the problem [Code: Problem Id] and explains what they're trying to do to solve it [Code: Communication]. B isn't a good brainstorm as it has no problem to solve [Code: Brainstorming] |

***Jatin***

Jatin also worked on the brainstorming topic in the same class as Amina. He was familiar with the comparison process and began working on the comparisons even before the teacher had fully given instructions to the class. He took a very short time working on the three comparisons he was presented. Two aspects stood out about the comparisons that Jatin made. First, in the case where he spoke about a range of justifications, his written comment remained brief and these details were lost. Second, his comments, in general, were brief. Despite the brevity of his comments and approach, following the LbE session Jatin was able to articulate several of the patterns he had seen in good-quality images, suggesting some impact of the experience.

We speculate two potential factors related to the quick pace of comparison that Jatin maintained. First, perhaps proficiency with the process was influential and allowed him to make the comparisons readily. Second, a pattern in the types of codes applied also suggests that the comparisons were focused on surface-level characteristics of the images (e.g., clarity of the image, coded as communication) or natural parts of the LbE process (e.g., describing the relative advantage between two images, coded as evaluation). These two codes—communication and evaluation—were by far the most frequent codes applied to Jatin's data. The simplicity of these comments may have facilitated the quick pace of comparison because they did not require much exploration or sense-making of the individual artifacts.

***Javier***

The final student we observed, Javier, took a moderate amount of time to compare the artifacts in his session. The focus of LbE prepared by his teacher was evaluating project management tactics used by design teams, specifically, evaluating Gantt charts created by previous design teams. The session required six to seven comparisons per student. As Javier began the TA process, he expressed initial discomfort and uncertainty about speaking out loud, saying he wasn't sure it would work because he was normally quiet. Despite this, with some prompting to begin speaking about what he saw in the artifacts, he continued to verbalize without issue.

Javier's approach for thinking in each comparison began by reading through each of the images carefully. Much of this verbalization was not necessarily coded, since it was seen as more of a sense-making process than an evaluative one. The nature of the artifacts may lend itself to this type of scrutiny. Because the Gantt charts are a visually dense representation of design work, longer time might be required to understand each example and then compare the differences between them. Given the visual character of the LbE session, the types of comments made by Javier included comments frequently coded as completeness, organization, or the design process steps included. However, this aligns with the intention of Gantt charts and could be reasonably expected by this type of session.

Javier's written comments were brief—while he understood the process, his responses were concise and tended to focus on one or elements. Because very few comments were made overall, we wondered whether Javier's case illustrates a broader concern about the quality of reasoning. Rather than seeing wide-ranging thinking and then limited written responses, our

judgment of Javier was that the initial level of evaluative thinking was low. Subsequently, with few details discovered when thinking, these would be easier to retain and more likely to be included in written work (an explanation for the harmony between the different forms of Javier's evaluation).

## Themes

Following the individual analysis, we discussed patterns and evidence across the student's experiences. These details helped us to arrive at three themes related to students' evaluative thinking process and the LbE process, which warrant further discussion and research. The themes identified were 1) familiarity with the evaluation process, 2) influence of teacher interactions and prompts, and 3) dimensions of filtering between thinking and written evaluation.

### *Familiarity With the Evaluation Process*

The observed students' comfort with the comparison process allowed them to express their thoughts freely and continuously. The fluency of thinking does suggest cognitive engagement from LbE, and affirms the usefulness of TA in this context. None of the students struggled with the LbE process or the software platform used. Nor did they struggle to translate their thinking into writing. However, in most cases students went through the comparisons quickly. As mentioned previously, this pace may have proven to be a limit in the quality of evaluative thinking. We are not privy to the established norms around the LbE process in each of these classes, but the rapid pace may have been established in earlier experiences with the process, despite its intent to foster deeper thinking, or the education climate. One other effect that we noticed about the process is that as students became familiar with the artifacts, their explanations tended to become less detailed. This may be related to activation of students' working memory and retention of the nature of the examples they have already seen.

### *Influence of Teacher Interactions and Prompts*

In this experience we observed several vectors whereby the teacher might influence students' thinking and writing. Teachers play an important role in influencing students (Bartholomew et al., 2023) and may be able to prompt deeper thinking, however, these verbal responses may not always translate into effective written comments by the students. As mentioned, teachers had participated in LbE sessions with their classes before, meaning that additional context and habits may be tacit in students' approach to thinking and writing here. Teachers directly introduced the session to students, focusing their engagement and reasoning during the LbE sessions. Although we did not analyze the class-wide discussion phase of LbE for data, we did observe this portion of the experience and affirm the importance of this discussion in reinforcing students' understanding. This may be a forum where students' thinking is further refined.

The discussion about teacher influence would be insufficient if it did not address the teachers' approach to setting up the LbE sessions for their class, including the intended learning outcome, holistic question, types of images shown, and the number of comparisons to be made. We observed a striking difference in the quality of comparisons when there was a large quantity required; when there were few comparisons required, students had the opportunity to be more

elaborate in their comparisons. Different types of images lead to different codes, as we had hypothesized in previous work (Jackson et al., 2022). For example, images about Gantt charts mainly produced codes on organization or completeness. In some of the sessions we noticed that images were included with low resolution, and in some cases broken links were included. In these cases, the natural reaction for students was to revert to surface-level comparisons rather than take effort to interpret what they were seeing. While we envisioned a possible teaching moment related to the broken links (i.e., reinforcing the implications of students turning in a wrong link for their own submission), this was not taken by the teacher and these items were confusing for students.

### *Cognitive Filtering in the Transition from Verbal to Written Expression*

In this experience, we observed that students' real-time verbalization allowed them to explore ideas more fluently than their written reflections. Across the students, the length of their verbal explanation was greater than what was written. This indicates that TA provides a unique space for exploratory thinking, whereas demands of writing may strain this process. In most cases that we observed, the filtering associated with writing was a refinement or condensing, and the focus of evaluations differed minimally.

Having expected to see a difference in the character of thinking and writing, that they were similar was a surprise to members of our team. However, in further discussion, we have posited several reasonable explanations. As indicated, the session logistics seem to provide a direct influence on students' thinking. The expectations students held for the session, and the ways they interpreted the number of comparisons required, both may have shaped the depth that they used in evaluation. Another explanation relates to the novice–expert continuum (Ericsson et al., 2007), and what may be expected of students' evaluative thinking in the 9th grade (Crismond & Adams, 2012). It may be possible that the threshold of evaluative thinking we observed here is expected for the cognitive development of these students. Finally, considering the relatively low level of evaluative thinking in the TA phase of the process, it may be the case that barriers between the thinking and writing phases are not yet experienced by these students. We see that when students are able to produce thinking focused on a small set of criteria, there is a high likelihood that they carry these criteria forward into writing. As students evaluative thinking, capacity increases, and the breath of their TA increases, we may see a departure from this trend.

## Conclusions

The opportunity to observe these classrooms, with attention specifically on students' thinking and writing has revealed opportunities to support students in the design process. Students' reasoning during design is an important indicator of their development as a designer (Crismond & Adams, 2012). Considering the discrete milestones of the LbE experience, and students' navigation from thinking to writing, highlights the need for timely interventions in student thinking. From the beginning of the process, the expectations set for students, thinking provide an influence. Setting clear expectations for the depth of evaluation, and the time afforded for the experience may help students be more comfortable to linger on this activity. We would suggest that LbE sessions use a limited number of comparisons to allow this depth.

Setting expectations about the written comments directly may also be helpful. Our observations here have provoked reflection that the written comments have most commonly been treated as a utility for researchers. Students may not be motivated to produce these written comments, and teachers may not be motivated to promote the writing phase, if it does not have direct utility for their learning. Prior research on writing from a theoretical view does suggest that it supports students' thinking (Graham, 2019), this is not as tangible as other forms of use. In one of the classes, we were inspired by the teacher showing written comments to students as a way to continue the conversation about what mattered in the upcoming design context.

Another opportunity that we imagine to deepen students' initial evaluative thinking is to pair students and invite them to think out loud together about the artifacts that they observe. This partnership might overcome some of the individual difficulties that students faced (e.g., applying the theory of the Zone of Proximal Development, Driscoll, 2005). Observing only a few students may inherit limitations related to their personalities (e.g., introversion or language difficulties) that influence our interpretation of students' evaluative thinking. However, treating the activity as a paired experience, may even work to overcome these individual limitations.

**References**


Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp 89–195). Academic Press.

Atman, C. J., Adams, R. S., Cardella, M. E., Turns, J., Mosborg, S., & Saleem, J. (2007). Engineering Design Processes: A comparison of students and expert practitioners. *Journal of Engineering Education*, *96*(4), 359–379. https://doi.org/10.1002/j.2168-9830.2007.tb00945.x

Atman, C. J., Chimka, J. R., Bursic, K. M., & Nachtmann, H. L. (1999). A comparison of freshman and senior engineering design processes. *Design Studies, 20*(2), 131–152. https://doi.org/10.1016/S0142-694X(98)00031-3

Baddeley, A. D. (2000). Short-term and working memory. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 77–92). Oxford University Press.

Bartholomew, S. R., Mentzer, N., & Jackson, A. (2023). Lessons from Dilbert: Clarifying design expectations. *Technology and Engineering Education, 1*(1), 7–13. https://www.iteea.org/journals

Bartholomew, S. R., Ruesch, E. Y., Hartell, E., & Strimel, G. J. (2019). Identifying design values across countries through adaptive comparative judgment. *International Journal of Technology and Design Education, 30*(2), 321–347. https://doi.org/10.1007/s10798-019-09506-8

Bartholomew, S. R., Barnum, E., Jackson, A., Mentzer, N., & Allen, J. (2023). The impact of teacher preferences in learning by evaluating. In S. Davies, M. McLain, A. Hardy, & D. Morrison-Love (Eds.), *Proceedings of the PATT40: The 40th International Pupils' Attitudes Towards Technology Research Conference*. https://doi.org/10.24377/PATT40.2023

Berninger, V. W., Garcia, N. P., & Abbott, R. D. (2009). Multiple processes that matter in writing instruction and assessment. In G. A. Troia (Ed.), *Instruction and assessment for struggling writers: Evidence-based practices* (pp. 15–50). Guilford Press.

Berninger, V. W., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Curtin, G., Hawkins, J. M., & Graham, S. (2002). Teaching spelling and composition alone and together: Implications for the simple view of writing. *Journal of Educational Psychology, 94*(2), 291–304. https://doi.org/10.1037/0022-0663.94.2.291

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219-243.

Cross, N. (2006). *Designerly ways of knowing*. Springer.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches.* Sage Publications.

Crismond, D. P., & Adams, R. S. (2012). The informed design teaching and learning matrix. *Journal of Engineering Education, 101*(4), 738–797. https://doi.org/10.1002/j.2168-9830.2012.tb01127.x

Driscoll, M. P. (2005). *Psychology of learning for instruction* (3rd ed.). Pearson.

Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership, 43*(2), 44-48.

Ericsson, K. A., Prietula, M. J., & Cokely, E. T. (2007, Jul-Aug). The making of an expert. *Harvard Business Review, 85*(7-8), 114–121, 193. https://www.ncbi.nlm.nih.gov/pubmed/17642130

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. MIT Press. https://doi.org/10.7551/mitpress/5657.001.0001

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365–387. https://doi.org/10.2307/356600

Goel, V., & Pirolli, P. (1992). The structure of design problem spaces. *Cognitive Science, 16*(3), 395–429. https://doi.org/10.1207/s15516709cog1603_3

Graham, S. (2019). Changing how writing is taught. *Review of Research in Education, 43*(1), 277–303. https://doi.org/10.3102/0091732X18821125

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3-30). Lawrence Erlbaum Associates.

Hayes, J. R., Flower, L., Schriver, K., Stratman, J. F., & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics: Reading, writing, and language processing* (pp. 176–240). Cambridge Press.

Jackson, A., Bartholomew, S., & Mentzer, N. (2022, Nov 17-18). *Priming the design process: Activating and characterizing students' critical thinking in design* [Paper presentation]. 108th Mississippi Valley Technology Teacher Education Conference and the 60th Southeastern Technology Education Conference, Nashville, TN.

Jackson, A., Mentzer, N., Bartholomew, S. R., Lee, W., Yauney, J. M., Thorne, S., & Bayah, D. (2023). Learning by evaluating (LbE): Engaging students in evaluation as a pedagogical strategy to improve design thinking. *Proceedings of the 2023 ASEE Annual Conference & Exposition*. https://peer.asee.org/42934

Kelley, T. R. (2011). Engineer's notebook – A design assessment tool. *Technology and Engineering Teacher, 70*(7), 30–35.

Kellogg R. T. (2001). Long-term working memory in text production. *Memory & Cognition, 29*(1), 43–52. https://doi.org/10.3758/bf03195739

Kim, Y. S. G., & Graham, S. (2021). Expanding the Direct and Indirect Effects Model of Writing (DIEW): Reading–writing relations, and dynamic relations as a function of measurement/dimensions of written composition. *Journal of Educational Psychology, 113*(4), 627-648. https://doi.org/10.1037/edu0000466

Kim, Y. S. G., & Park, S. H. (2019). Unpacking pathways using the direct and indirect effects model of writing (DIEW) and the contributions of higher order cognitive skills to writing. *Reading and Writing, 32*, 1319-1343. https://doi.org/10.1007/s11145-018-9913-y

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

Lee, W., Mentzer, N., Clevenger, A., Jackson, A., & Bartholomew, S. (2024). Integration of learning by evaluating (LbE) within the 5e instructional model in engineering design education. *Proceedings of the 2024 ASEE Annual Conference & Exposition.*

Mentzer, N., Becker, K., & Sutton, M. (2015). Engineering design thinking: High school students' performance and knowledge. *Journal of Engineering Education, 104*(4), 417–432. https://doi.org/10.1002/jee.20105

Mentzer, N., Lee, W., Jackson, A., & Bartholomew, S. (2023). Learning by evaluating (LbE): Promoting meaningful reasoning in the context of engineering design thinking using

adaptive comparative judgment (ACJ). *International Journal of Technology and Design Education, 34*, 1145–1169. https://doi.org/10.1007/s10798-023-09853-7

Mentzer, N., Lee, W., Jackson, A., & Bartholomew, S. (2024). *Defining and evaluating argumentation quality in the context of design thinking: Using high school students' design critiques from foundational engineering courses* [Manuscript submitted for publication].

Nicol, D. J., and Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218. https://doi.org/10.1080/03075070600572090

Pugalee, D. K. (2004). A Comparison of Verbal and Written Descriptions of Students' Problem Solving Processes. *Educational Studies in Mathematics 55*, 27–47. https://doi.org/10.1023/B:EDUC.0000017666.11367.c7

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144. https://doi.org/10.1007/bf00117714

Sampson, V., & Clark, D. B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education, 92*(3), 447–472. https://doi.org/10.1002/sce.20276

Swanson, H. L., & Berninger, V. W. (1996). Individual differences in children's working memory and writing skill. *Journal of Experimental Child Psychology, 63*(2), 358–385. https://doi.org/10.1006/jecp.1996.0054

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285.

Tashakkori, A., & Teddlie, C. (2010). *SAGE Handbook of Mixed Methods in Social & Behavioral Research*. SAGE Publications. https://doi.org/10.4135/9781506335193

van Hattum-Janssen, N., & Lourenço, J. M. (2006b). Explicitness of criteria in peer assessment processes for first-year engineering students. *European Journal of Engineering Education, 31*(6), 683–691. https://doi.org/10.1080/03043790600911779

Wilson-Lopez, A., Strong, A. R., Hartman, C. M., Garlick, J., Washburn, K. H., Minichiello, A., Weingart, S., & Acosta-Feliz, J. (2020). A systematic review of argumentation related to the engineering-designed world. *Journal of Engineering Education, 109*(2), 281–306. https://doi.org/10.1002/jee.20318