



A systematic literature review of student evaluation of peer exemplars and implications for design, Technology, and Engineering Learning

Scott Thorne¹ · Nathan Mentzer¹ · Scott Bartholomew² · Greg J. Strimel¹

Accepted: 11 December 2023 / Published online: 16 January 2024
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract

In design courses, reviewing how others have solved design problems or completed projects is common practice and often encouraged by educators. Using student work as examples can provide context for assessment criteria and help students approach new design problems. While studies have explored the use of exemplars in various disciplines, little research has focused on which exemplars to use (e.g., high-quality, low-quality) in design, technology, and engineering fields. To address this gap, researchers conducted a literature review of 33 articles on exemplar use in secondary and post-secondary education. The analysis revealed nine themes related to exemplar use and their impact on student learning, including (1) Clarity of instruction, (2) Learner focus, (3) Motivation for learning, (4) Student reflection on learning, (5) Building student self-efficacy, (6) Identifying instructional challenges, (7) Providing contrasting cases, (8) The relationship between exemplar quality and student work quality, and (9) Raising the bar for learning outcomes. Findings suggest that simply providing an exemplar is not enough and that the selection of an exemplar can have positive or negative impacts on student motivation, understanding, and application. Carefully selecting exemplars and engaging in dialogue with students can help them identify expectations, recognize quality work, and identify potential misconceptions. These findings have implications for those involved in design, technology, and engineering education. Educators can use these findings to guide their selection of exemplars and engage students in meaningful dialogue to aid their learning. Researchers can also use these findings to further investigate the use of exemplars in these fields.

Keywords Student evaluation of exemplars · Selection of exemplars · Design-based learning · Student work quality · Building student self-efficacy · Motivation for learning

✉ Scott Thorne
sthorne@purdue.edu

¹ Purdue University, West Lafayette, IN, USA

² Brigham Young University, Provo, UT, USA

Introduction

One important factor for student success, across all levels of education, is the ability to critically evaluate their work's quality in alignment with their instructor's expectations (Amulya, 2004; Anderson & Freiberg, 1995; McMillan & Hearn, 2008). In an effort to aid students to be independent self-regulating learners, rubrics and criteria are often provided as part of design, technology, and engineering courses. These tools can communicate expectations with students and allow students to evaluate their own work, including open-ended design tasks, against a set of standards (Andrade, 2000). The type of rubric employed (analytic or holistic) is determined by the scope, purpose, and value attributed by researchers or practitioners in its context of application.

However, research has shown that relying solely on rubrics is inadequate for guiding student learning as “verbal descriptions are always to some degree vague or fuzzy. A fuzzy standard cannot, therefore, be defined into existence” (Sadler, 1987, p. 202). For this reason, Sadler (1987), recommended to choose and present students with samples or peer exemplars of relevant coursework that align with the desired criteria in context, this is because the assessment process involving exemplars and rubrics can be intertwined to mutually reinforce each other. While literature has established that selecting exemplars that highlight important criteria is beneficial, little definition exists regarding how the quality of the exemplars impact learning.

This review encompasses diverse studies on the use of peer exemplars across various fields including business, design, and science (Bouwer et al., 2018; Carless & Chan, 2017; Knight et al., 2019; Kulkarni et al., 2014) to synthesize how exemplar use impacts learning. Participants in these studies are often presented with full text documents in which they evaluate the exemplar against a rubric to determine the criteria (clarity), define what makes it ‘good’ (quality), and identify misconceptions when attempting a similar task (calibration). This systematic literature review relies on secondary data sources to develop an informed stance about how the attributes of exemplar selection influence learning. This is done to advance both design-based learning and the learning of design.

Research purpose

The purpose of this systematized review was to search, identify, and synthesize themes from existing literature on the role of the evaluation of peer exemplars—samples of previous similar work—in the learning experience. This was approached by identifying relevant literature in the field, selective linking to expand the pool of relevant literature, and identifying themes from these studies for a systematized search. The systematized review was guided by the following questions:

- (1) What are the potential influences of reviewing peer exemplars on student learning?
- (2) How does the selection of peer exemplars influence the learning experience?

Given the intention to apply findings from this research into design, technology, and engineering education, initial searches focused on publications in the related educational journals or conference proceedings. Specific methods for conducting and refining such a

search are described below. Following the search protocol descriptions, themes identified from existing research emerged to address the above questions.

Method

Strategies used to refine a set of search terms specific to the above research questions were patterned from recommendations in conducting a systematic literature review by Borrego et al. (2014). The use of peer exemplars and evaluation in education has been explored in multiple studies and disciplines, however, the intentional selection of peer exemplars, or which exemplars to show students, remains unanswered and unexplored. Therefore, as suggested by Peticrew and Roberts (2008), a systematic literature review was chosen to evaluate existing research in a “transparent, methodical, and reproducible” manner (Borrego et al., 2014, p. 46).

Initial investigations to establish search criteria consisted of 23 articles selected based on teaching methods that rely on students interacting with and evaluating examples such as gallery walks, math thinking, worked examples, and visual thinking strategies. However, these did not provide data to answer the research questions. The primary focus of each was instead the student performance after embedding said strategy into the curriculum than the strategy itself. One article that not only began to address the research questions, but cited several relevant studies was “Why and how educators use exemplars” (Hawe et al., 2021). Through selective linking, four more high quality articles were identified and analyzed for key themes to start a systematic search (Carless et al., 2018; Carless & Chan, 2017; Hawe et al., 2019; Newlyn, 2013). A synthesis of these key articles as well as the previous 23 articles that focused on instructional strategies related to learning through the evaluation of peer exemplars resulted in the establishment of keywords for a targeted search (see Appendix 1).

Journals related to design, education, technology, technology education, engineering, and engineering education were identified for their relevance on how to identify effective peer exemplars. For this reason, the databases EBSCOhost (ERIC), Scopus, and Elsevier (Engineering Village) were chosen as they index field specific conference proceedings including the American Society for Engineering Education (ASEE) and IEEE Frontiers in Education (FIE). Initial numbers of results from these databases were quite high. EBSCOhost (ERIC) returned 910, Scopus returned the most at 451,504, and Elsevier (Engineering Village) returned 14,422 articles. These results were narrowed by several factors including full text availability, peer reviewed, only articles from the past 10 years (2012–2022), and articles in relevant subject areas (i.e., engineering and math). This brought total search results down to 259 articles in EBSCOhost (ERIC), 382 in Scopus, and 231 in Elsevier (Engineering Village).

The following criteria were established to prioritize studies with an empirical nature that explored the use of exemplars and their influence on student learning:

- (1) The methods used in each study included an evaluation of samples, examples, or exemplars which may have been current student work, previous student work or from sources external to the classroom (e.g., professional work) where the central concept is consciously looking such that students are engaged in review or evaluation of the artifact to consider its perceived merits, scope and limitations.
- (2) The focus of the article was on supporting student learning.

- Search results including the title, author, and abstract from each database were exported into individual spreadsheets in an Excel workbook. After eliminating duplicates, the number of unique articles was reduced to 872. Next, a face validity check was performed on the article titles, further narrowing the selection down to 158 articles. These 158 articles were further refined by assessing their content validity, which involved an analysis of both the abstracts and methods sections. This rigorous process resulted in seven articles for an in-depth analysis, namely those authored by Dixon et al. (2020), Gonczi et al. (2017), Hawe and Dixon (2017), Headley and Pittson (2020), Kean (2012), Kulkarni et al. (2014), and Rawson and Dunlosky (2016). The same inclusion and exclusion criteria used during the face validity assessment of article titles were consistently applied to maintain methodological rigor, relevance to exemplars, and contextual appropriateness.

To ensure the reliability of inclusion and exclusion decisions, frequently cited literature that was not included in the initial search results were reexamined. It was noted that while previously identified authors from the five articles focused on exemplars were represented in the dataset (Hawe, Carless, Chan, Newlyn), none of the articles themselves were present. Further, frequently cited authors, such as Carless, had no articles represented in the search results. At an attempt to be more inclusive, the original five articles on peer exemplars were made into a word frequency cloud using NVivo software (see Fig. 1). A new keyword search was developed using words unique to these articles, and passed again into EBSCOhost (ERIC), Scopus, and Elsevier (Engineering Village) (see Appendix 2). This additional step of checking for frequently cited publications that were not present in the first search query may help improve the reliability of the study.

Fig. 1 Exemplar word cloud

past 10 years (2012–2022). This further reduced the number of results to 45 in EBSCOhost (ERIC), 14 in Scopus, and 3 in Elsevier (Engineering Village).

Search results from each database were exported, including the title, author, and abstract, into individual spreadsheets in an Excel workbook. Once combined, with duplicates removed, the total number of unique articles was 59. Again, the same inclusion and exclusion criteria were utilized on each title for face validity, reducing the total number down to 36 articles. Each of the remaining articles were further checked for content validity by assessing the abstract and methods section, reducing the total number to 24 articles. These 24 articles focused on exemplars were combined with the seven articles from the initial search strategy and the five from the initial exemplar articles, and, after removing duplicates, was reduced to 33 articles for final synthesis (see Figs. 2 and 3).

Data analysis

After finalizing a list of articles, each study was coded in NVivo 12 using open-coding procedures (Merriam & Tisdell, 2015). Although typically associated with qualitative research, this method was adapted for the systematic review as it offers a structured process for identifying any valuable data segments and aids in the development of categories and themes. Initially, a comprehensive reading of each article was conducted to establish familiarity. During this initial reading, codes were generated based on recurring keywords, phrases, and themes within the article, while detailed notes were recorded in a linked memo. Codes from each additional study were developed in much the same way using the constant-comparison technique to identify commonalities and deviations from initial codes (Glaser, 1965). As coding progressed and discernable patterns began to emerge, axial coding was used to reorganize and establish a hierarchy to create overarching themes (Charmaz, 2006). Following the identification of themes and a thorough understanding of the codes, each article underwent a secondary reading to enhance data familiarity and confirm alignment with the overarching themes.

Findings

The research questions focused on the influence of reviewing peer exemplars on student learning, and how the selection of exemplars influences the student learning experience. During the thorough examination of documents within NVivo through open coding techniques (Saldaña, 2013), overarching themes emerged from the existing research. Rather than strictly aligning these themes with specific questions, it became evident that they were intricately interconnected. As a result, we present the following comprehensive nine themes, each of which will be reviewed in detail:

- (1) Clarity of instruction.
- (2) Learner focus.
- (3) Motivation for learning.
- (4) Student reflection on learning.
- (5) Building student self-efficacy.
- (6) Identifying instructional challenges.
- (7) Providing contrasting cases.
- (8) The relation between exemplar quality and student work quality.

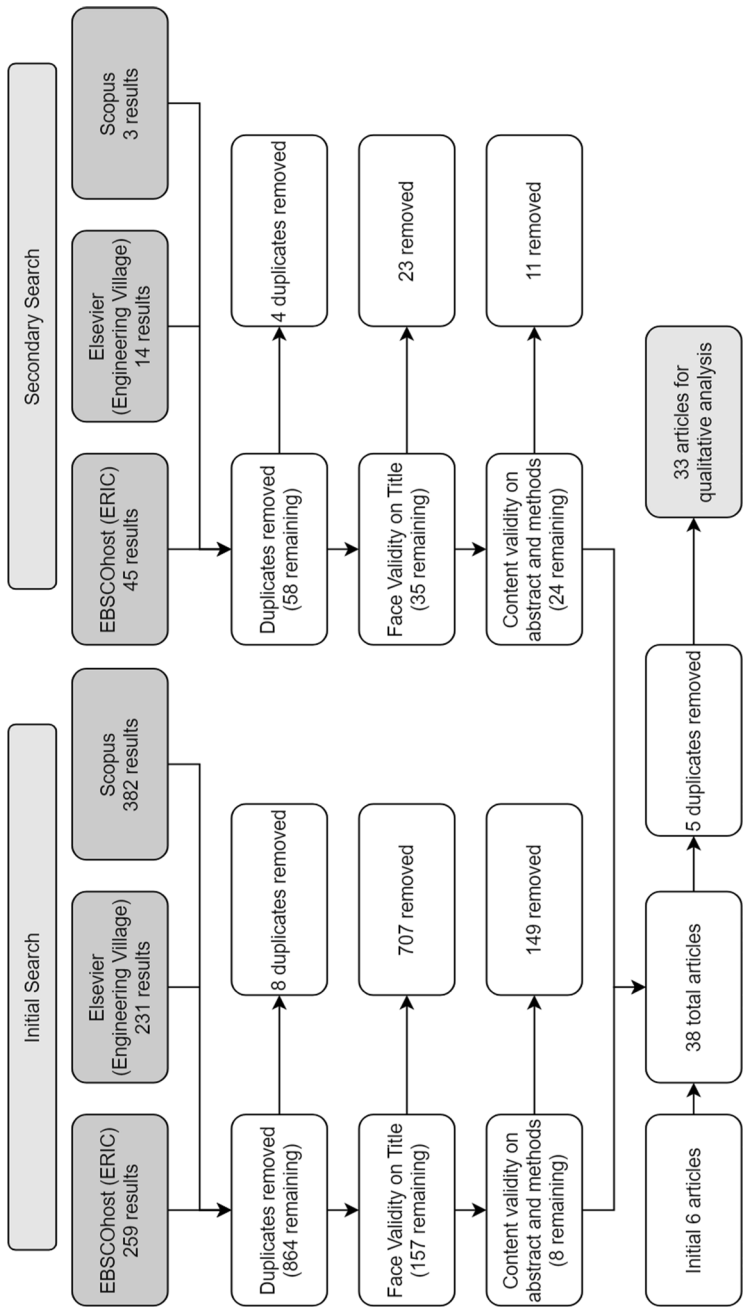


Fig. 2 Flowchart of search results by selection criteria

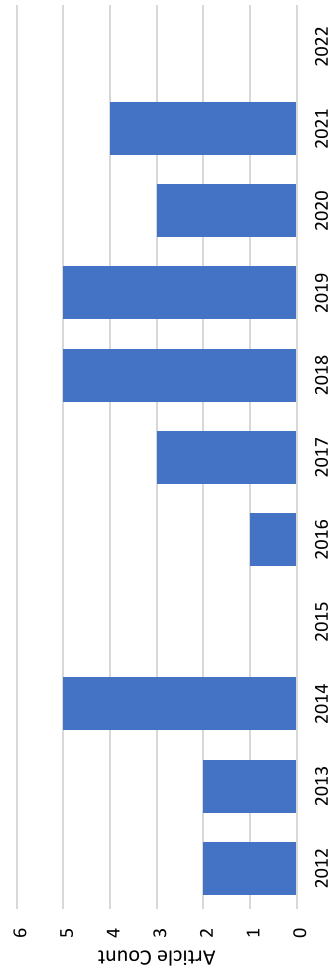


Fig. 3 Number of articles in final selection by year from 2012 – 2022. *Note* Refer to the inclusion and exclusion criteria for the selection of these articles

(9) Raising the bar for learning outcomes.

In the following sections, each theme will be described in the context of students evaluating peer exemplars as a learning experience. Citations and excerpts from the final articles selected for review, including qualitative comments from participants in their studies, are included below to contextualize and provide credibility to the findings.

Theme 1: clarity of instruction

Expanding and defining project criteria emerged as a major theme for the role of peer exemplar evaluation in enhancing the learning experience. This approach enables teachers to provide clarity and helps students recognize the quality or lack of quality of their work more effectively than other mechanisms (Sadler, 2002). Establishing clarity early in the process allows for both teachers and students to be in alignment on assignment expectations (Chong, 2021; Grainger et al., 2018; Hawe & Dixon, 2017; Headley & Pittson, 2020; Hendry et al., 2012; Yucel et al., 2014) as assessment criteria was sometimes be viewed as too dense or difficult to comprehend (Bell et al., 2013; Carless et al., 2018). These points were bolstered by quotes from students such as, “*I would not have known what to do if we had not done this. It gave me an accurate expectation*” (Yucel et al., 2014, p. 976). A student from another study had a similar sentiment, “*By you giving us those examples, we can determine what you deem to be the best answer, so we can kind of structure our own letter to match that, or improve on that*” (Hendry et al., 2012, p. 153).

Establishing clarity for expectations at the start of a project can allow for students to establish forethought as they begin to plan for short and long-term goals (Zimmerman, 2000), especially as they are able to better visualize requirements and expectations (Aitken & Thompson, 2018; Bell et al., 2013; Grainger et al., 2018; Tam, 2021). Providing transparency through peer exemplars was received very positively by students (Carless et al., 2018; Chong, 2021; Headley & Pittson, 2020; Yucel et al., 2014). One student even described how the clarity in expectations helped bring understanding to grade distributions, making for a more enjoyable experience even when underperforming:

Since this process I find that I am more aware now of what's expected of me in assessments and the marking criteria and how markers approach them is now more clear. I'm also no longer disappointed with a mark as I have some idea of what I'll receive. (Kearney & Perkins, 2014, p. 5)

Clarity for learning, a major theme across multiple studies, may also be at the center of each of the themes across all study contexts. Having clarity in a project allows for focus, provides the opportunity for motivation, and enables reflection. Without fully understanding expectations, students are not likely to develop self-efficacy, identify challenges, or know where to start if attempting to raise the bar. It is the umbrella term that is in the background of each of the following themes.

Theme 2: learner focus

When attempting a project, students are often lost on where and how to start, even with a well written project description and detailed rubrics (Grainger et al., 2018). In fact, too detailed of rubrics can confuse students as they tend to present information in abstract terms (Bouwer et al., 2018; Sadler, 1989, 2002) and the more standards and criteria are

broken down, “the harder it is to make the bits work together as a whole” (Sadler, 2007, p. 389). Hendry and Tomitsch (2014) stated that when left to their own devices, students often ignore rubrics and criteria until they have nearly completed their task. One student elaborates “*you don’t even read the criteria of what you’re making. You... go out, make it and think, ‘Oh my God, this is the best ever’ [but miss an important criterion]. It happens.*” (Hendry & Tomitsch, 2014, p. 342). Evaluation of peer exemplars shifts the focus back to what is being assessed, encouraging students to check their assumptions against an agreed upon standard (Hendry & Tomitsch, 2014; Kearney & Perkins, 2014; Knight et al., 2019).

Peer exemplars help to address this by showing what the criteria look like in a final product, allowing students to take a step back and see the bigger picture (Aitken & Thompson, 2018). In one study, researchers noted the shift in attention that when students evaluated documents side by side, explaining “Compared to the criteria condition, students in the comparative judgment condition [holistically evaluating two exemplars side-by-side] focused relatively more on aspects that were related to the content and structure of the text, and less so on aspects that were related to grammar and vocabulary.” (Bouwer et al., 2018). Using peer exemplars can help broaden students’ focus and enable them to differentiate between various levels of achievement in areas that are not covered by rubrics or task guidelines due to spatial constraints, such as formatting and layout. (Grainger et al., 2018).

Students too found it useful to see what the teacher might focus on when evaluating an assignment (Hendry & Jukic, 2014; Hendry & Tomitsch, 2014; Kean, 2012), as one stated, “*It is always good to see how a teacher will evaluate your work. Some teachers have different styles and look for different things. Listening to the teacher [talk about peer exemplars] made me understand what to focus on.*” (Hendry & Tomitsch, 2014, p. 341). After evaluation of peer exemplars, students not only started to recognize patterns of expectations (Kean, 2012; Yucel et al., 2014), they were also more aware of the current state of their work and found themselves more receptive to feedback (Kearney & Perkins, 2014; Tam, 2021).

Theme 3: motivation for learning

The evaluation of authentic peer exemplars, those created by students, can be a source of motivation for starting a project (Grainger et al., 2018; Headley & Pittson, 2020) as it shows students not only what it might look like, but that it can be done (Dixon et al., 2020; Hendry & Jukic, 2014). One postgraduate student shared “*Going through it with [the teacher] at the end and seeing what [the teacher] gave marks for did make you think, ‘I actually – I could do that, I could do that’*” (Hendry & Jukic, 2014, p. 9). Other students noted that the process of making evaluations of peer exemplars was rather exciting and highly engaging (Kearney & Perkins, 2014), with some looking forward to seeing if they were in alignment with their teacher:

The first stage [marking] is making you aware of your own thoughts and what’s going on in your own head and how you’re judging things and then, when [the teacher explains], there’s this big reveal kind of thing. You wanted to know, the suspense was killing you. (Hendry & Tomitsch, 2014, p. 342)

While peer exemplars can help to motivate students, changing the quality of exemplars too far in either direction (i.e., much higher or lower quality) can act as a demotivating factor. Sharing work that is of very high quality, especially when only showing high quality work, can cause students to doubt their capabilities (Hawe & Dixon, 2017)

and undermine confidence (Dixon et al., 2020), but surprisingly, sharing low quality examples can cause similar reactions, especially with authentic peer exemplars. In one study, students felt anxious seeing that past students had performed poorly (Hendry & Jukic, 2014), while in a separate study, students described examples of students who had failed as “unhelpful, demotivating, and even distracting” (Grainger et al., 2018, p. 5). Recommendations to combat this include sharing a range of peer exemplars, either all of higher quality, or a wider range that goes no lower than ‘passing’ (Bouwer et al., 2018; Dixon et al., 2020; Hawe et al., 2019; Hawe & Dixon, 2017; Hendry et al., 2012). When applied correctly, however, peer exemplars of the appropriate level of complexity can motivate and inspire students “to ‘match or beat’ the quality and standard” (Bell et al., 2013, p. 772).

Theme 4: student reflection on learning

While not frequently cited, the act of metacognition when interpreting tacit knowledge is essential to the student learning experience. Evaluation helps to make this tacit knowledge explicit, as the process helps trigger a reflective process as considering the kinds of feedback they might provide others aids in transfer to self-assessment (Bell et al., 2013; Tam, 2021; Yucel et al., 2014).

To help foster this, it is strongly encouraged, by nearly every study in this review focusing on the use of evaluating peer exemplars as an instructional method, that teachers model their thinking and engage students in dialogue after assessment (Aitken & Thompson, 2018; Bell et al., 2013; Bouwer et al., 2018; Carless et al., 2018; Carless & Boud, 2018; Carless & Chan, 2017; Chong, 2019, 2021; Dixon et al., 2020; Grainger et al., 2018; Hawe et al., 2019, 2021; Hawe & Dixon, 2017; Hendry et al., 2012; Hendry & Jukic, 2014; Hendry & Tomitsch, 2014; Johnston et al., 2021; Kean, 2012; Tam, 2021; To & Liu, 2018). One of the key reasons for this, is that students often under- or overestimate the work of others, as one student describes “*We were marking it really low ... Or really high. We were thinking, ‘It’s a 7, or it’s a 6’, and [they’d] go, ‘It’s a 4’, or ‘It’s a 3’. We were like, ‘Wow!’*” (Hendry et al., 2012, p. 155). Other studies found that evaluating peer exemplars made students more attentive, asking questions about what makes a good answer. An undergraduate student elaborated:

It made you pay attention more, so – then because you’re curious too – yeah, because you can’t assess how [the teacher’s] marking unless you work out how you would mark it.... So you need to do that, otherwise it wouldn’t sink in – I don’t think it would sink in as much, what [the teacher] would say. (Hendry & Jukic, 2014, p. 7)

Reflection when evaluating can also be enhanced through the use of rubrics or explicitly stated criteria either immediately before or during the marking session. One student reflected on his assumptions and knowledge about topics such as ethics when marking an exemplar:

The marking criteria was useful as it provided avenues for reflection. For me it highlighted the elements of ethical perspectives and reasoning that should have drawn from the case. The marking criteria made me critically reflect on whether individual assumptions and knowledge about ethics were as strong and consistent as I previously thought. (Bell et al., 2013, p. 777)

Encouraging reflection during the process of evaluation can help students to internalize tacit knowledge, calibrate their understanding, and develop their own understanding of expectations.

Theme 5: building student self-efficacy

Student self-efficacy is a significant predictor of success and achievement for students (Zimmerman & Schunk, 2007), and can be fostered through the evaluation of peer exemplars. The act of analyzing and evaluating peer exemplars has been shown to reduce some of the anxiety associated with the uncertainty of assessment standards and requirements (Carless & Boud, 2018; Carless & Chan, 2017; Hawe et al., 2019, 2021; Yucel et al., 2014). In fact, by evaluating multiple peer exemplars, students internalize a wider range of approaches to the assessment (Chong, 2019), which may help address concerns for stifling creativity and design fixation.

Students engaging in exemplar evaluation, across multiple studies, indicated feeling more confident to work independently on challenging assignments (Bell et al., 2013; Dixon et al., 2020; Hawe et al., 2019, 2021; Headley & Pittson, 2020; Hendry & Tomitsch, 2014). In a focus group with other professors who utilize peer exemplars at Harper Adams University, one professor shared that the evaluation of exemplars “*gives students the confidence to work out how to do the assignment themselves without constant recourse to tutor to ask whether it is right*” (Headley & Pittson, 2020, p. 23). This is reflected in Zimmerman’s (2000) theory of self-regulated learning, frequently cited in the literature around synthesis, which focuses on self-regulating behavior as a core component toward goal-completion.

Researchers further describe the observation of self-efficacy in the actions of their students after evaluating peer exemplars as taking ownership in their work and actively monitoring and regulating their progress in concluding remarks:

They were setting goals using the exemplars as benchmarks or targets; they were devising strategies to ensure they avoided plagiarism; they were managing resources (i.e., the various exemplars provided) by determining what exactly they were choosing to focus on, whether it was a FAILED exemplar or a quality exemplar or academic literacies. (Grainger et al., 2018, p. 8)

This, again, closely aligns with the establishment of appropriate actions, thoughts, feelings, and behaviors central to self-regulated learning (Zimmerman, 2000). Additionally, it seems that by seeing authentic peer exemplars, even of B and C level work, students who might doubt their abilities are inspired that they too can succeed (Dixon et al., 2020; Hawe et al., 2021; Hawe & Dixon, 2017).

Theme 6: identifying instructional challenges

Peer exemplars are carefully chosen samples that are representative of student work, and while they are used to clarify expectations, they may also be used to better visualize different dimensions of quality (Carless & Chan, 2017). While often associated with high quality work, it is common to provide student samples that highlight common mistakes and pitfalls that are made in assessments throughout the semester (Aitken & Thompson, 2018; Dixon et al., 2020; Grainger et al., 2018; Hawe et al., 2021; Headley & Pittson, 2020; Kean, 2012; Newlyn, 2013; Rawson & Dunlosky, 2016; Tam, 2021; Yucel et al., 2014).

In general, researchers felt that this was an opportunity for students to learn from the mistakes of others and while some take a passive role, providing an exemplar and letting students ask questions or come to their own conclusions (Bell et al., 2013; Kean, 2012; Kearney & Perkins, 2014; Newlyn, 2013; Rawson & Dunlosky, 2016), other researchers advocate for dialogue to help explain *why* students may score poorly (Aitken & Thompson, 2018; Dixon et al., 2020; Hawe et al., 2021; Tam, 2021). This strategy of dialogue has gained a following in math education. Understanding and walking students through the thought process behind mistakes to learn from them have inspired pedagogical strategies, such as “My Favorite No” (Teaching Channel, 2014), websites dedicated to cataloging wrong answers and how to address them (mathmistakes.org), book chapters (Seeley, 2014), and entire books (Anderson, 2017).

Other researchers methodologies utilized the experience of having students take a more active role by either providing a rating for the assessment (Aitken & Thompson, 2018; Headley & Pittson, 2020; Hendry & Tomitsch, 2014), or even correcting the mistakes they see in peer exemplars, offering suggestions on how it might be improved (Dixon et al., 2020; Hawe et al., 2019, 2021; Tam, 2021; Yucel et al., 2014).

Not all authors in the articles reviewed are in agreement with the practice of evaluating low quality peer exemplars, as some researchers state that it may draw too much focus on what not to do, rather than highlighting best practices. Hendry and Tomitsch (2014) explain their reasoning for only using ‘pass’, ‘credit’, and ‘high distinction’ peer exemplars with students, stating:

We believe that providing students with a ‘fail’ exemplar might be counterproductive because it could focus students on what not to do rather than what to strive for. It might also challenge students’ positive outlook and belief in their ability to be successful in their design project. (Hendry & Tomitsch, 2014, p. 344)

This statement is supported by previous work by Hendry and Tomitsch, finding that in courses where students were presented with low quality peer exemplars and the teacher emphasized errors and low quality, the students “felt less confident to attempt their assignment” (Hendry et al., 2012, p. 157). Other researchers acknowledge that the use of ‘fail’ level peer exemplars may be counterproductive, as they do not show what is needed to achieve (Aitken & Thompson, 2018), and may even demotivate students (Grainger et al., 2018). Interestingly, when low quality samples were absent from studies, students indicated that they wanted them included (Bell et al., 2013), with one student explaining how they “Find it useful, of how people have approached it and how that’s worked, so like not just shining examples but where people have gone wrong perhaps and how to avoid that” (Kean, 2012, p. 92). In summary, providing too low quality of peer exemplars is counterproductive but leaving them out entirely leaves students uncertain of how to fully interpret assignment criteria.

Theme 7: providing contrasting cases

While Schwartz’s (2018) did not meet our inclusion criteria, it offers valuable insights into how students can benefit from reviewing exemplars of varying complexity. Contrasting cases guide students to discern similarities and differences in exemplar qualities, promoting deeper understanding of learning objectives and preparing them for future knowledge transfer. Although direct comparisons of drastically different peer exemplars were limited in the studies cited (Bouwer et al., 2018; Chong, 2021; Hendry & Jukic,

Table 1 Comparing qualities of peer exemplars

Quality provided	References
Credit and above	Bell et al. (2013), Hendry & Tomitsch, (2014), Knight et al. (2019), Newlyn, (2013)
Average and high quality	Bouwer et al. (2018)
High, average, and low	Chong, (2019), Grainger et al., (2018), Hendry et al. (2012), Kean, (2012), Newlyn, (2013)
High and low	Chong, (2021), Hendry & Jukic, (2014), Yucel et al. (2014)
Unspecified range	Hawe & Dixon, (2017), Newlyn, (2013)

In “Providing exemplars in the learning environment: the case for and against” Newlyn (2013) cites several studies with varying methodological approaches

2014), many studies presented a range of student work to more thoroughly define rubric criteria. To illustrate this range of approaches, we have included a table (Table 1) summarizing the methods of comparison employed across studies. Referencing the quality of peer exemplars evaluated when reviewing conclusions of various researchers helps to contextualize the student experience. It is worth noting that while other studies included in this systematic review may have also specified the use of different quality peer exemplars, only those that explicitly tied student performance to contrasting quality are identified in the table and the subsequent discussion.

When presented with peer exemplars of contrasting quality, students begin to differentiate between average quality and high-quality work, and expectations of the teacher of what constitutes high quality work (Bouwer et al., 2018; Chong, 2019, 2021; Grainger et al., 2018; Hendry & Jukic, 2014; Kean, 2012). A common trend from these studies is that they all provide either two comparisons (average against high quality or high against low quality), or a range of three peer exemplars (high, average, and low). Studies listed as providing qualities of peer exemplars that are “credit and above” typically describe the quality of their exemplars as being closer together such as “pass, distinction, and high distinction” (Bell et al., 2013; Hendry & Tomitsch, 2014; Newlyn, 2013) or evaluations that are even closer together such as “pass, credit, distinction, and high distinction” (Knight et al., 2019). While it may be that the distinctions students make when exposed to a variety of quality levels was not the interest of researchers conducting the study, and thus not reported, it could also be that as items are closer on the spectrum of quality, it becomes more difficult to distinguish one from another, especially when evaluated against a criterion.

The use of contrasting cases did not always focus on quality of tasks, at times a variety of approaches were shared as well (Bell et al., 2013; Grainger et al., 2018; Hendry & Tomitsch, 2014). This too is important, as a common criticism of using peer exemplars is the suppression of creativity (Headley & Pittson, 2020; Hendry & Jukic, 2014; Newlyn, 2013). However, when presented with a variety of approaches, students expressed that seeing peer exemplars may have helped them to be *more* creative (Hawe et al., 2021; Tam, 2021). One student shared how the evaluation of peer exemplars enhanced their design experience:

Because we saw such a range of exemplars, like comparing the one where they had the cages, to the one where they did it in violet colours and moods and stuff. They were two completely different ideas. That doesn't say to me, I'm going to pick one

and copy it. It says, there are so many ideas, what's mine? (Hendry & Tomitsch, 2014, p. 343)

Another interest of researchers was that students not only recognized what makes something 'good', but that their own performance on assignments improved (Bouwer et al., 2018; Hawe & Dixon, 2017; Knight et al., 2019; Newlyn, 2013). One student explained their perceived benefit in an interview:

You can compare, because he (the teacher) gave us two passages. So, while you are comparing the two passages, you can consider why this one gets higher mark, or somehow, to learn the ... how to say ... to learn from this comparison, and to improve your writing skills. (Chong, 2021, p. 21)

While the Hawe and Dixon study (2017) did not specify the range of peer exemplars provided to students, the common thread for other studies indicated an increase in performance was the lack of a low-quality example. Further, Yucel (2014) found that students who participated in their program achieved significantly lower marks than those who did not. While there are other contributing factors to these findings including student buy-in, lack of teacher-led discussion, and dissatisfaction with peer review, contributions from other studies indicate that too low of peer exemplars can have a negative impact on student understanding, motivation, and application (Aitken & Thompson, 2018; Grainger et al., 2018; Hendry et al., 2012; Hendry & Tomitsch, 2014).

Theme 8: the relationship between exemplar quality and student work quality

When considering theories of how peer exemplars might influence student work, one theoretical underpinning was the concept of tacit knowledge acquisition or tacit knowing (Carless & Boud, 2018; Carless & Chan, 2017; Chong, 2019). That is to say that not all that we know can be verbalized, and there is much that is understood or implied without being stated, relying on the experience of observation, imitation, and practice as it is otherwise difficult to share one's thinking process (Polanyi, 1997).

Another set of theories frequently used were Vygotsky's sociocultural theory (SCT) and zone of proximal development (ZPD) (Carless et al., 2018; Chong, 2019, 2021; Tam, 2021; To & Liu, 2018). Sociocultural theory posits that learning takes place through social interactions with others, and is refined with the zone of proximal development, the space between what a learner can do independently and what they can do when collaborating with others or otherwise supported (Vygotsky, 1987). Accordingly, these studies emphasized a procedure where students would evaluate exemplars independently, with a peer, and through dialogue with their teacher.

The socio-cognitive work of Zimmerman (2000; 2007) was also used in multiple studies for the model of self-regulation (Grainger et al., 2018; Hawe et al., 2019, 2021). The model of self-regulation is a three-phase model consisting of forethought, performance, and self-reflection, with specific components to each phase, such as task analysis and self-motivational beliefs in the phase of forethought (Zimmerman, 2000). The act of evaluating the work of others was often linked to the task analysis, which leads to self-motivational beliefs as students realize that they are capable and progresses to performance where they can apply the insights, and finally to self-reflection to evaluate the degree of success in their attempt. Zimmerman's theory is built on work by Bandura which was also cited for his theory of self-efficacy in one study (Dixon et al., 2020), aligning with the first phase of forethought with regard to self-motivational beliefs. The theory of self-efficacy suggests

that the motivations of a student are driven by the confidence they have to complete a task (Bandura, 1977).

While only used in one study (Chong, 2021), Nelson's (2018) conceptual framework of evaluative judgement was used to describe how students focus on various parts of work when making evaluations. Evaluative judgement is described in terms of hard, soft, and dynamic. In hard evaluative judgement, students focus on the mechanics of an exemplar, such as punctuation and grammatical decisions. Soft evaluative judgement tends to focus on word choice and patterns of speech. However, dynamic evaluative judgement is related to creativity and how ideas are organized (Nelson, 2018). It is a higher level of thinking, and one that students are scaffolded to through dialogue with peers and their teacher-led discussions.

Multiple theories were used to justify how peer exemplars might transfer the knowledge of what makes an educational artifact 'good' to influence student performance, which are supported by evidence found in the research. Several studies have drawn correlations between using the evaluation of peer exemplars as learning strategy with improvements on assessments (Bell et al., 2013; Carless et al., 2018). One student voiced their experience in an interview, "*When we use exemplar, we would automatically compare our own articles with the exemplar, thinking about the differences between our articles and the exemplar, then I know what I should improve on*" (Chong, 2021, p. 21). Other researchers expressed similar observations, noting how students would use peer exemplars as a benchmark to make adjustments to their own work (Hawe et al., 2019; Headley & Pittson, 2020). Additionally, in a longitudinal study, it was found that students who evaluated peer exemplars outperformed their control group peers, but even a year after the initial study the participants were achieving significantly better results (Bell et al., 2013; Rawson & Dunlosky, 2016). Part of the value of evaluating peer exemplars is reflecting on other approaches, especially if students are experiencing difficulties with not knowing where to start. One graduate student explained in an interview:

It puts a structure in your head so when you see the question you're like, 'Okay.' You know how to break it down instead of before [when] you'd see the question but, 'Okay I've got all this knowledge in my head and I'm just going to blurt it out whereas in no logical order. (Hendry & Jukic, 2014, p. 6)

Other research warns that while students will eventually apply what they have seen from peer exemplars, results for some may come more slowly than others (Kean, 2012). This study describes how students first pushed back when asked to evaluate their peers, and after the second week the majority of students "failed to show they could identify 'quality' in others' work or produce it in their own" (Kean, 2012, p. 89). However, after repeated cycles and dialogic feedback students were able to identify and correctly apply criteria by the fifth week. This iterative process is consistent with other research on evaluation of peers and exemplars (Bell et al., 2013; Carless & Boud, 2018; Chong, 2019; Knight et al., 2019).

Theme 9: raising the bar for learning outcomes

There seems to be some consensus in the advantages to engaging students by evaluating high quality work. In designing studies, excellent peer exemplars were occasionally chosen "on the grounds that students might have more to learn from high-quality work" (Carless & Chan, 2017, pp. 932–933). This may influence a student's own work as "students can

sharpen appreciation of how quality is manifested, enhance their capacities to make sound academic judgments, and draw appropriate inferences for actions” (Carless & Boud, 2018, p. 1321). When selecting appropriate peer exemplars and students are actively engaging, there is a belief that students will push themselves to do better (Bell et al., 2013; Dixon et al., 2020; Hawe et al., 2019; Hawe & Dixon, 2017; Hendry & Tomitsch, 2014). ‘Appropriate’ peer exemplars, are thought to be ones that are high, but not too far above a student’s current capabilities, as one researcher explains:

If the intent is to build robust efficacy beliefs for all students, use of a range of exemplars from across the grade range is recommended. We believe using those from the mid to low A range can encourage the more capable students who like to ‘rise to the challenge’. (Dixon et al., 2020, p. 465)

The risk of presenting too high quality work is that students not only neglect to try to raise the bar, but they may also start to question their ability to successfully complete the task entirely (Dixon et al., 2020; Hawe et al., 2019, 2021; Hawe & Dixon, 2017). Dixon expands on this idea as they warn:

Given the fragility of academic self-efficacy, the need for a delicate balance between building confidence and overwhelming students should be taken into account. Only selecting exemplars that exemplify the highest standards of achievement such as those illustrative of an A plus piece of work may be inspiring for some students, but for others such work may undermine confidence. (2020, p. 465)

Student interviews from several studies support the notion that high quality peer exemplars motivate students to perform at a higher level. After evaluating samples from a variety of levels, one student stated that they felt “*I can do better than that*” and “*felt I could achieve*” (Hawe & Dixon, 2017, p. 1185). From an interview in a different study, other students shared similar sentiments “*when seeing the quality of work [that got] higher marks, it encourages me to try and produce the same kind of work*” and “[...] you kind of want to aim higher [...]” (Hawe et al., 2019, p. 37). This extended past just the quality of work to the creativity of work as well as a student reflected on how peer exemplars pushed them to raise the bar in their design work, sharing “*It made you want to work harder to... try and be more creative*” (Hendry & Tomitsch, 2014, p. 342).

Discussions and recommendations for practice

According to literature presented in this review, there can be many benefits to using exemplars in design-focused coursework, including building self-efficacy, improving the quality of student work, and motivating students. Importantly, research supports that the careful selection and consistent use of exemplars contribute to the overall goal of providing clarity on instructor expectations beyond what is included in the task guidelines (Sadler, 1987, 1989, 2002, 2007). Peer exemplars can provide an aid to students when differentiating between a variety of levels of achievement through tangible examples. Through the synthesis of themes from this study, the careful selection of exemplars can promote productive classroom dialogue (themes 4, 6, 8). It can help students identify expectations (themes 1, 2, 4, 6, 7) while also serving to identify potential misconceptions leading to improved student understanding (themes 1, 4, 7). Moreover, it aids in student motivation (themes 1, 3, 8, 9) and application (themes 3, 7, 8) in engineering and design education.

When considering how the quality of peer exemplars influence student performance, literature from this review supports that in general, high-quality exemplars are most effective. This exposes students to effective problem-solving strategies and design solutions, while inspiring students and increasing confidence in their ability to succeed. Exemplars of too low of quality may confuse students and lead to misconceptions on expectations and how to achieve success. These low-quality examples may even bring about feelings of discouragement and low self-efficacy among students. Finally, the quality of exemplars may have an impact on dialogic discussions. Higher-quality exemplars provide an opportunity for students and teachers to have a clear standard with which to evaluate work, making it easier to provide constructive feedback and identify areas of improvement. This is in contrast to lower-quality exemplars, which may make discussions challenging beyond identifying areas of improvement.

This collection of literature also addressed the role of exemplar quality on student decision making and self-efficacy in several ways. First, high-quality exemplars can provide students with a clear understanding of what constitutes a successful design solution, increasing confidence in their ability to produce similar outcomes. Second, seeing examples of well-designed solutions can inspire students to think more creatively and take risks in their own work. On the other hand, low-quality exemplars may lead to lower expectations of what is possible, negatively affecting confidence in their abilities and motivation to engage in the design process.

When identifying high-quality exemplars for peer evaluation, literature in this review cautioned against selecting exemplars too far above the students' current level of achievement. Using peer exemplars that are of extremely high-quality may create unrealistic expectations and a sense of intimidation for students. This may leave students feeling overwhelmed as they may not be able to reach the level of quality presented, leading to reduced self-efficacy and motivation, negatively impacting their performance in the related coursework. Additionally, providing too high-quality exemplars may not provide enough variety for students to differentiate levels of achievement for acceptable work. This could lead to frustration and potential disengagement from the course.

Beyond the scope of initial research questions, articles in this literature review overwhelmingly supported dialogic feedback following the evaluation of peer exemplars to address student misconceptions. Teachers should facilitate discussions around exemplars to highlight strengths, areas of improvement, and opportunities for transfer to help students apply what they have seen to their work. This can help scaffold student learning, especially as it pertains to evaluation criteria and how that criterion aligns with learning objectives for the course. By doing so, teachers can help students develop a more accurate understanding of the quality of their own work and improve their decision making and self-efficacy in the course. Overall, peer exemplars can significantly impact student decision making and self-efficacy in an engineering and design course, highlighting the importance of carefully selecting and using exemplars as a pedagogical tool.

Research recommendations

Authors cited in this study focused their research on the fields of foreign language, teacher education, law, first-year college experience courses, and other areas outside of design or engineering curriculum. Accordingly, for those wishing to expand this systematic literature review, it is recommended to expand search terms and include other education research databases.

Additionally, results of the systematized literature review revealed both a scarcity of research on the purposeful selection of exemplars, and a limited focus on artifacts emphasizing writing in short answer and essay format, presented as a final product. Additional research around the influence of evaluation on student learning, and the purposeful selection of sketches, images, short statements, and video as exemplars for evaluation may build on current findings. This holds numerous implications for STEM education, as it assists students in visualizing abstract design problems, recognizing diverse approaches to problem-solving, facilitating comparisons among potential solutions, and synthesizing information to innovate new designs.

Another recommended area of future research is if the selection of exemplars is as important as how they are used in the classroom. Common to nearly all articles identified for this study is the emphasis on the role of the teacher and dialogic strategies to frame the evaluation session before starting and to debrief with a partner and in large groups after. Often, students arrived at erroneous conclusions or fail to reach consensus on the quality of an exemplar when discussing with a peer or small group, necessitating a debriefing session. While the evaluation of exemplars remained the primary focus of the articles, in seven of the nine themes there was a repeated emphasis on how much the role of the teacher influences such an activity (themes 1, 2, 3, 4, 6, 7, 8).

Conclusion

The purpose of this research was to investigate foundational literature supporting the selection of peer exemplars to support student achievement in design, technology, and engineering related fields. The systematized search narrowed initial search results from 467,223 to 33 articles focused on secondary and post-secondary evaluative experiences over the past 10 years. It was found that research on the effectiveness of peer exemplars is primarily focused on written student samples for large projects, and the purposeful selection of peer exemplars is limited to two to five samples for a given class. Research on peer exemplars reflects a gap in studies focused on the variety of sample quality in peer exemplars for evaluation, especially in STEM-related fields.

In analyzing the influence of peer exemplars on the impact on student's learning, nine themes were identified to address the research questions, specifically, (1) How the quality of exemplars influences student performance, (2) How comparing drastically different examples influences decision making, and (3) The effect of providing students with lower quality work. In addressing student performance, it was found that quality exemplars assist students with clarity and understanding of the task and what makes something 'good', motivate students in their own tasks, and facilitate knowledge transfer to enhance performance.

A variety of ways that instructors and teachers use peer exemplars in the classroom were discovered such as to showcase a wide range of student work and using drastically different examples. It was found that by providing a spectrum of samples, students being able to differentiate between average and high-quality work (Bouwer et al., 2018; Chong, 2019, 2021; Grainger et al., 2018; Hendry & Jukic, 2014; Kean, 2012). Tangential to this, studies found that multiple peer exemplars counteract claims that providing students with examples hinders creativity. Findings indicate that when multiple peer exemplars are used it instead acts to inspire students into more divergent ideas when problem solving (Hawe et al., 2021; Hendry & Tomitsch, 2014; Tam, 2021).

Providing lower quality examples likewise helped students to identify challenges and misconceptions by learning from the mistakes of others. However, counterintuitive to teacher expectations, effects of using peer exemplars that are too high or too low of quality were shown to demotivate students and have a negative impact on understanding the assessment (Dixon et al., 2020; Hawe et al., 2019, 2021; Hawe & Dixon, 2017). As many of the studies cited contained qualitative aspects, multiple quotes from student reflections on the use of peer exemplars were introduced to strengthen the conclusions drawn from this paper and the claims made by researchers.

Limitations

This systematic review has some methodological limitations that need to be acknowledged. Firstly, as the search and coding were conducted by the first author, there is potential for bias in the search and selection process. While efforts were made to conduct a comprehensive search using multiple databases and search terms, the inclusion and exclusion criteria were set by a single author. This may have led to the exclusion of relevant studies or the inclusion of studies that do not meet the criteria.

An additional limitation revolves around the possibility of overlooking certain studies. Despite the efforts to conduct a thorough search, it is possible that pertinent studies might have been missed, particularly those that were not present in the searched databases, accidentally screened out, or not published in English. This may have resulted in a skewed sample of studies.

The act of synthesizing existing literature into thematic patterns inherently carries limitations. It is possible that the nuanced details of how exemplars are applied in various scenarios, as well as the true impact of student learning, may not have been fully captured. To mitigate this, several steps were taken. A diverse sample selection was employed, broadening the spectrum of examples and their patterns of use. An in-depth analysis was conducted to identify recurring patterns or trends related to the use of examples in different contexts. Quotations and citations were used whenever possible to provide contextual examples to add credibility and depth to the analysis.

Finally, the generalizability of the findings may be limited due to the specific inclusion and exclusion criteria used in this study. The themes identified may not be representative of all studies on the topic, and the findings may not be applicable to other contexts or populations.

Appendix 1

Initial Search Query (Conducted 7/20/2022).

#	Query
S1	"exemplars" OR "samples" OR "worked examples" OR "examples" OR "model answer" OR "Gallery walk" OR "Design critique" OR "Peer critique" OR "Formative peer assessment" OR "Math thinking" OR "Learning by Evaluation" OR "LbE" OR "Peer learning"
S2	"self-efficacy" OR "self efficacy" OR "self-confidence" OR "self confidence" OR "confidence" OR "efficacy" OR "concept of quality" OR "evaluative knowledge"
S3	"secondary school" OR "high school" OR "secondary education" OR "postsecondary education" OR "college" OR "university"
S4	"selection" OR "quality" OR "choice"

Appendix 2

Additional Search Query (Conducted 7/22/2022).

#	Query
S1	“assessment” OR “marking” OR “criteria” OR “standards”
S2	“discussion” OR “task” OR “understanding”
S3	“exemplars”
S4	“education”

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Competing Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

- Aitken, A., & Thompson, D. G. (2018). Using software to engage design students in academic writing. *International Journal of Technology and Design Education*, 28(3), 885–898. <https://doi.org/10.1007/s10798-017-9413-4>
- Amulya, J. (2004). *What is reflective practice?* Center for Reflective Community Practice at the Massachusetts Institute of Technology.
- Anderson, N. (2017). *What's right about wrong answers: Learning from math mistakes, grades 4–5*. Stenhouse Publishers.
- Anderson, J. B., & Freiberg, H. J. (1995). Using self-assessment as a reflective tool to enhance the student teaching experience. *Teacher Education Quarterly*, 22(1), 77–91.
- Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13–19.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bell, A., Mladenovic, R., & Price, M. (2013). Students' perceptions of the usefulness of marking guides, grade descriptors and annotated exemplars. *Assessment & Evaluation in Higher Education*, 38(7), 769–788. <https://doi.org/10.1080/02602938.2012.714738>
- Borrego, M., Foster, M. J., & Froyd, J. E. (2014). Systematic literature reviews in engineering education and other developing Interdisciplinary fields. *Journal of Engineering Education*, 103(1), 45–76. <https://doi.org/10.1002/jee.20038>
- Bouwer, R., Lesterhuis, M., Bonne, P., & De Maeyer, S. (2018). Applying criteria to examples or learning by comparison: Effects on students' evaluative judgment and performance in writing. In *Frontiers in education* (Vol. 3, p. 86). Frontiers Media SA. <https://doi.org/10.3389/feduc.2018.00086>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Carless, D., & Chan, K. K. H. (2017). Managing dialogic use of exemplars. *Assessment & Evaluation in Higher Education*, 42(6), 930–941. <https://doi.org/10.1080/02602938.2016.1211246>
- Carless, D., Chan, K. K. H., To, J., Lo, M., & Barrett, E. (2018). Developing students' capacities for evaluative judgement through analysing exemplars. *Developing Evaluative Judgement in Higher Education: Assessment for Knowing and Producing Quality Work*, 108–116.

- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. SAGE.
- Chong, S. W. (2019). The use of exemplars in English writing classrooms: From theory to practice. *Assessment & Evaluation in Higher Education*, 44(5), 748–763. <https://doi.org/10.1080/02602938.2018.1535051>
- Chong, S. W. (2021). University students' perceptions towards using exemplars dialogically to develop evaluative judgement: The case of a high-stakes language test. *Asian-Pacific Journal of Second and Foreign Language Education*, 6(1), 12. <https://doi.org/10.1186/s40862-021-00115-4>
- Dixon, H., Hawe, E., & Hamilton, R. (2020). The case for using exemplars to develop academic self-efficacy. *Assessment & Evaluation in Higher Education*, 45(3), 460–471. <https://doi.org/10.1080/02602938.2019.1666084>
- Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social Problems*, 12(4), 436–445. <https://doi.org/10.2307/798843>
- Goncz, A. L., Bergman, B. G., Huntoon, J., Allen, R., McIntyre, B., Turner, S., Davis, J., & Handler, R. (2017). Decision matrices: Tools to enhance middle school engineering instruction. *Science Activities*, 54(1), 8–17. <https://doi.org/10.1080/00368121.2016.1264922>
- Grainger, P. R., Heck, D., & Carey, M. D. (2018). Are assessment exemplars perceived to support self-regulated learning in teacher education?. *Frontiers in Education*, 3, 60.
- Hawe, E., & Dixon, H. (2017). Assessment for learning: A catalyst for student self-regulation. *Assessment & Evaluation in Higher Education*, 42(8), 1181–1192. <https://doi.org/10.1080/02602938.2016.1236360>
- Hawe, E., Lightfoot, U., & Dixon, H. (2019). First-year students working with exemplars: Promoting self-efficacy, self-monitoring and self-regulation. *Journal of Further and Higher Education*, 43(1), 30–44. <https://doi.org/10.1080/0309877X.2017.1349894>
- Hawe, E., Dixon, H., & Hamilton, R. (2021). Why and how educators use exemplars. *Journal of University Teaching and Learning Practice*. <https://doi.org/10.53761/1.18.3.10>
- Headley, J., & Pittson, H. (2020). Bottom up! Establishing a University Wide community of practice for exemplars and exploration of its affect in enhancing the quality of student assessment. *Practitioner Research in Higher Education*, 13(1), 18–26.
- Hendry, G. D., & Jukic, K. (2014). Learning about the quality of work that teachers expect: Students' perceptions of exemplar marking versus teacher explanation. *Journal of University Teaching and Learning Practice*, 11(2), 5.
- Hendry, G. D., & Tomitsch, M. (2014). Implementing an exemplar-based approach in an interaction design subject: Enhancing students' awareness of the need to be creative. *International Journal of Technology and Design Education*, 24(3), 337–348. <https://doi.org/10.1007/s10798-013-9256-6>
- Hendry, G. D., Armstrong, S., & Bromberger, N. (2012). Implementing standards-based assessment effectively: Incorporating discussion of exemplars into classroom teaching. *Assessment & Evaluation in Higher Education*, 37(2), 149–161. <https://doi.org/10.1080/02602938.2010.515014>
- Johnston, E., Burleigh, C., Rasmussen, X., Turner, P., Valentine, D., & Bailey, L. (2021). Multimedia open educational resource materials for teaching-online diversity and leadership: Aligning bloom's taxonomy and studio habits of mind. *Higher Learning Research Communications*, 11(2), 40–67.
- Kean, J. (2012). Show AND tell: Using peer assessment and exemplars to help students understand quality in assessment. *Practitioner Research in Higher Education*, 6(2), 2.
- Kearney, S. P., & Perkins, T. (2014). Engaging students through assessment: The Success and limitations of the ASPAL (authentic self and peer assessment for learning) Model. *Journal of University Teaching and Learning Practice*, 11(3), 2.
- Knight, S., Leigh, A., Davila, Y. C., Martin, L. J., & Krix, D. W. (2019). Calibrating assessment literacy through benchmarking tasks. *Assessment & Evaluation in Higher Education*, 44(8), 1121–1132. <https://doi.org/10.1080/02602938.2019.1570483>
- Kulkarni, C. E., Socher, R., Bernstein, M. S., & Klemmer, S. R. (2014). Scaling short-answer grading by combining peer assessment with algorithmic scoring. In *Proceedings of the First ACM conference on learning @ scale conference* (pp. 99–108). <https://doi.org/10.1145/2556325.2566238>
- McMillan, J. H., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons*, 87(1), 40–49.
- Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research* (4th ed.). Wiley.
- Nelson, R. (2018). *Barriers to the cultivation of evaluative judgement: A critical and historical perspective. Developing evaluative judgement in Higher Education*. Routledge.
- Newlyn, D. (2013). Providing exemplars in the learning environment: The case for and against. *Universal Journal of Educational Research*, 1(1), 26–32.
- Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. Wiley.

- Polanyi, M. (1997). The Tacit Dimension. In *Knowledge in Organizations*. Routledge.
- Rawson, K. A., & Dunlosky, J. (2016). How effective is example generation for learning declarative concepts? *Educational Psychology Review*, 28(3), 649–672. eric.
- Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education*, 13(2), 191–209. <https://doi.org/10.1080/0305498870130207>
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144. <https://doi.org/10.1007/BF00117714>
- Sadler, D. R. (2002). Ah!... so that's 'quality'. In *Assessment*. Routledge.
- Sadler, D. R. (2007). Perils in the meticulous specification of goals and assessment criteria. *Assessment in Education: Principles Policy & Practice*, 14(3), 387–392. <https://doi.org/10.1080/09695940701592097>
- Saldaña, J. (2013). *The Coding Manual for Qualitative Researchers* (Second). SAGE Publications Inc.
- Schwartz, D. (2018). *Designing contrasting cases for inductive learning*. Human Sciences and Technologies Advanced Research Institute. <https://hstar.stanford.edu/research/designing-contrasting-cases-for-inductive-learning/>
- Seeley, C. L. (2014). Chapter 8: Oops! The value of mistakes, failures, and wrong turns. In *Smarter than we think: More messages about Math, Teaching, and learning in the 21st Century: A Resource for teachers, leaders, Policy makers, and families*. Math Solutions.
- Tam, A. C. F. (2021). Undergraduate students' perceptions of and responses to exemplar-based dialogic feedback. *Assessment & Evaluation in Higher Education*, 46(2), 269–285. <https://doi.org/10.1080/02602938.2020.1772957>
- Teaching Channel. (2014). *My Favorite No (Uncut)*. Teaching Channel. <https://learn.teachingchannel.com/video/my-favorite-no-complete-lesson>
- To, J., & Liu, Y. (2018). Using peer and teacher-student exemplar dialogues to unpack assessment standards: Challenges and possibilities. *Assessment & Evaluation in Higher Education*, 43(3), 449–460. <https://doi.org/10.1080/02602938.2017.1356907>
- Vygotsky, L. S. (1987). *The collected works of L.S. Vygotsky: The fundamentals of defectology*. Springer Science and Business Media.
- Yucel, R., Bird, F. L., Young, J., & Blanksby, T. (2014). The road to self-assessment: Exemplar marking before peer review develops first-year students' capacity to judge the quality of a scientific report. *Assessment & Evaluation in Higher Education*, 39(8), 971–986. <https://doi.org/10.1080/02602938.2014.880400>
- Zimmerman, B. J. (2000). Chapter 2 - Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 13–39). Academic Press. <https://doi.org/10.1016/B978-012109890-2/50031-7>
- Zimmerman, B. J., & Schunk, D. H. (2007). *Motivation: An essential dimension of self-regulated learning. Motivation and self-regulated learning*. Routledge.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.