A systematic review of the educational uses and effects of exemplars

Jessica To[a], Ernesto Panadero[b c] & David Carless[d]

Author Note

[a]National Institute of Education, Nanyang Technological University, Singapore.

[b]Facultad de Educación y Deporte, Universidad de Deusto, Bilbao, España.

[c]IKERBASQUE, Basque Foundation for Science, Bilbao, Spain.

[d]Faculty of Education, The University of Hong Kong, Hong Kong.

Recommended citation:

To, J., Panadero, E., & Carless, D. (2021). A systematic review of the educational uses and effects of exemplars. *Assessment & Evaluation in Higher Education,* 1-16. https://doi.org/10.1080/02602938.2021.2011134

Correspondence concerning this manuscript should be addressed to: Jessica To.
National Institute of Education, Nanyang Technological University, Singapore.
E-mail: jessica.to@nie.edu.sg

Abstract

The analysis of exemplars of different quality is a potentially powerful tool in enabling students to understand assessment expectations and appreciate academic standards. Through a systematic review methodology, this paper synthesises exemplar-based research designs, exemplar implementation and the educational effects of exemplars. The review of 40 empirical exemplar-based studies indicated frequent use of post-intervention evaluation designs. The integrated use of exemplars and strategies such as rubrics, peer or self-assessment was a major pedagogic feature to enhance students' understanding of standards and aid their generation of internal feedback for self-monitoring. There was evidence suggesting that the combined use of exemplars with these strategies could advance students' academic performance and development of academic self-regulation. The review results set future directions for exemplar-based studies: quantitative research would benefit from controlled manipulation of variables to disentangle the effects of exemplars and those of other strategies; qualitative research could profitably use self-reflective diaries, think aloud protocols or classroom observations to develop deeper understandings of exemplar implementation. The key pedagogical implication recommends students' production of an assignment draft prior to exposure of exemplars so that they can compare their own work with exemplars and then make improvements independently of teachers.

Keywords: Exemplars; research designs; exemplar implementation; educational effects

A systematic review of the educational uses and effects of exemplars

Introduction

The widespread adoption of standards-based assessment in university highlights the need for clear communication of assessment standards to students (Rust, Price, and O'Donovan 2003; Smyth and Carless 2021). It can, however, be challenging for teachers to clarify assessment standards to students due to the opaqueness of rubrics and criteria (Hawe and Dixon 2017). Exemplars are one of the pedagogical tools that carries significant potential in helping students grasp tacit standards and are defined as "key examples chosen so as to be typical of designated levels of quality or competence" (Sadler 2005, 192).

Exemplar use is gaining momentum in higher education for four main reasons. First, the concrete nature of exemplars aids students in interpreting criteria and assessment standards and understanding teachers' expectations of performance (e.g. Hendry, Armstrong, and Bromberger 2012). Second, exchanging views about how exemplars are appraised enables students to develop academic judgement and assessment literacy (e.g. Knight et al. 2019). Third, exemplars support the development of academic literacy because students could learn how to communicate disciplinary ideas through unpacking the features of different genres (e.g. Wu 2019). Fourth, exemplars have potential to promote self-regulated learning (SRL) (e.g. Hawe, Lightfoot, and Dixon 2019). SRL can be operationalised when students make reflective comparisons between exemplars and their own work-in-progress and then make improvements to their work before final submission (Carless 2020; Nicol 2021).

Notwithstanding their potential value, academics sometimes express concerns about students' copying from exemplars (Handley and Williams 2011), inhibition of creativity (Hendry and Tomitsch 2014) and instrumental learning (Torrance 2007). These concerns could potentially be reduced if teachers have a more sophisticated understanding of

exemplar-based practices. To clarify what constitutes productive exemplar use and research practices, this paper synthesises empirical research designs, the implementation of exemplars and their educational effects. The contributions to the scholarship of exemplars lie in critiquing the methodological designs of exemplar-based studies and making recommendations to improve research designs and exemplar-related pedagogy.

Knowns and unknowns of exemplar use

To set the scene for the review, we identify the knowns and unknowns of exemplar implementation and the educational effects of exemplars.

There is no clearly agreed recipe for exemplar implementation, but the broad consensus is that students need active engagement with exemplars to develop their understanding of assessment standards (O'Donovan, Price, and Rust 2008). Active engagement involves students in making and explaining judgements of exemplars and producing insights for self-monitoring (To and Carless 2016). This could be achieved in various pedagogical activities: grading workshops (Rust et al. 2003; Wimshurst and Manning 2013); discussion of exemplar grades with peers and teachers (Hendry, White, and Herbert 2016; Tam 2021); and analysis of the strengths and weaknesses of exemplars (Macbeth 2010; Carless and Chan 2017). All these activities signify dialogic use of exemplars which refers to the knowledge construction process when students actively participate in the discussion of exemplars (Carless and Chan 2017; To and Liu 2018). Although this pedagogic feature is prominent in exemplar-based practices, it mainly focuses on students' cognitive development through socialisation. It is unclear how this feature could raise students' metacognitive awareness and encourage academic self-regulation.

Another important pedagogic feature is the use of rubrics in exemplar-based practices. Rubrics are a useful complement to exemplars because rubrics can facilitate dialogues between students and teachers to clarify meanings of criteria (Panadero and Jonsson 2020).

Conversely, exemplars help to address the problem of indeterminacy of criteria (Sadler 2009). The combined use of rubrics and exemplars takes three directions. The first provides students with exemplars and rubrics without explanation in the pre-assessment stage (Lipnevich et al. 2014). This requires students to interpret both exemplars and rubrics independently to self-monitor performance. The second involves the deconstruction of rubrics prior to exemplar grading or analysis (Jones et al. 2017). While this method supports students' application of criteria to make judgements, Sadler (2009) critiques it for limiting students' judgement-making to a list of predetermined criteria. The third invites students to co-construct rubrics and criteria with peers after exemplar analysis (Ayalon and Wilkie 2020). Criterion co-construction is useful in deepening students' engagement with standards and their awareness of assessment criteria (Bacchus et al. 2020).

Concerning the educational effects of exemplars, one of the obvious effects is increased clarity of assessment standards. Since students can acquire a better understanding of task requirements and quality, they generally welcome exposure to exemplars in the pre-assessment stage (Hendry et al. 2012; Blair et al. 2014). This positive perception is particularly common among first-year undergraduates who require support in becoming acculturated into the university's academic community (Yucel et al. 2014).

Another effect is students' changes in academic performance after exemplar use. The first focal point of this issue is whether exemplar use could improve academic performance. Rust et al. (2003) observed an improvement in performance after students participated in exemplar grading, discussion and self-assessment of draft, but Smith et al. (2013) found no differences in students' performance after dialogic use of exemplars. This raises the question whether the inconsistent results were due to different evaluation methods, students' academic abilities or other contextual factors. Importantly, the field lacks a systematic exploration of this issue. Our review aims to investigate it further.

The second focal point of this issue is in what ways exemplars advance performance. There have been reports on students' higher level of confidence and self-efficacy for task engagement after they have a clearer understanding of assessment requirements (e.g. To and Carless 2016; Hawe et al. 2019). This points to self-motivation beliefs, one of the subprocesses in the forethought phase of Zimmerman's (2000) SRL model. It is unknown how exemplar use could facilitate students' SRL development in the other two phases (performance and self-reflection).

In short, we have identified the knowns and unknowns of exemplar use in higher education. While dialogic use of exemplars and complementary use of rubrics and exemplars are common characteristics, it is unclear whether exemplars improve academic performance and SRL. Furthermore, there has been scant discussion of the evaluation methods used in exemplars research. To advance the development of exemplar-based research, we aim to evaluate the quality of exemplar-based studies and make recommendations to improve research designs and exemplar implementation. The guiding research questions are as follows:

RQ1. What are the research designs of exemplar-based studies?

RQ2. How are exemplar-based practices implemented?

RQ3. What are the educational effects of exemplar use?

Method

We employed the systematic review methodology because it helped us scientifically identify pertinent studies to answer the research questions, synthesise key findings and indicate future directions for topic development (Garg, Hackham, and Tonelli 2008). Its benefits include increasing the transparency of search process and avoiding researcher bias in literature selection (Mallet et al. 2012).

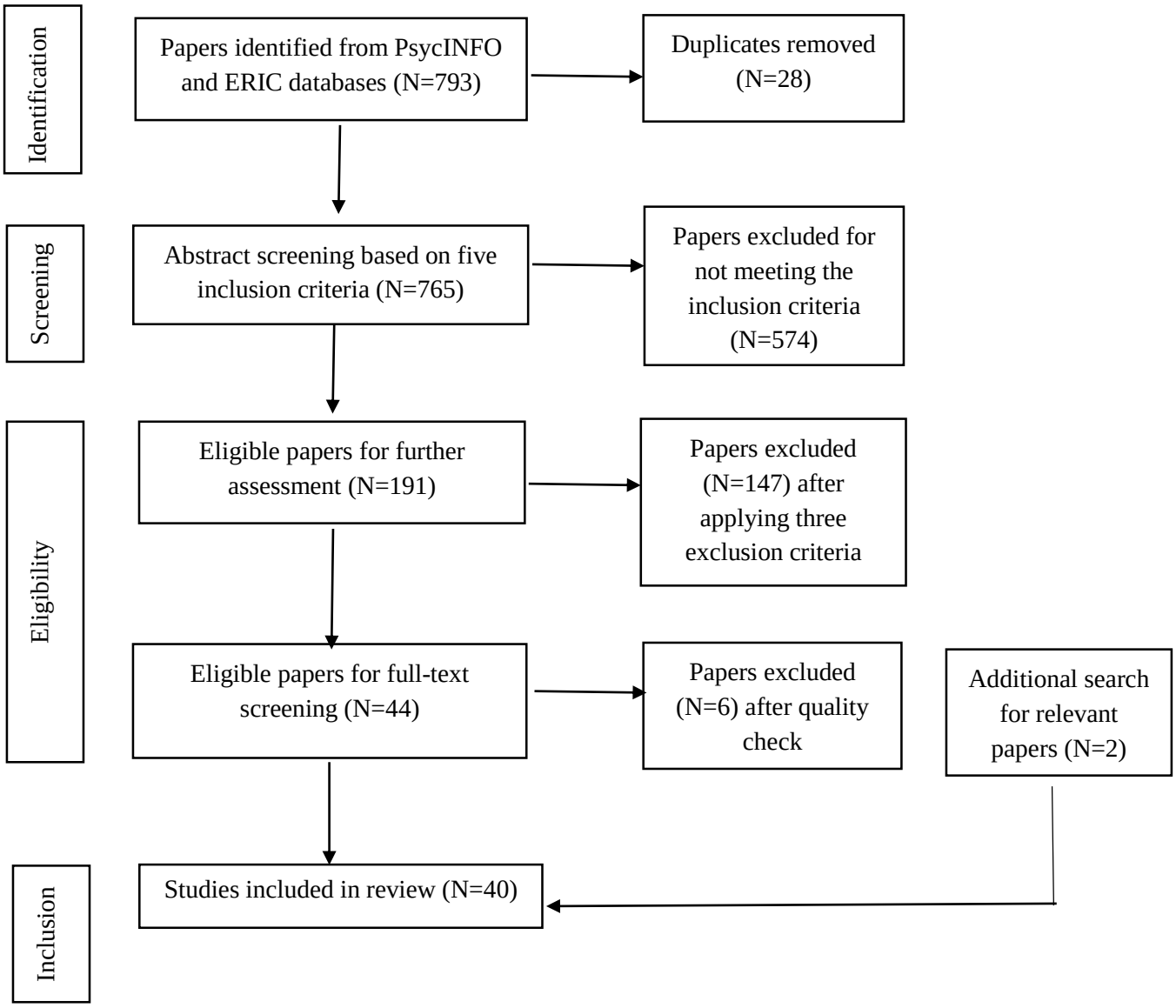
Selection of studies

The search was conducted in July 2020 in ERIC and PsycINFO databases, with the following keyword combinations for the fields of title and abstract: exemplar / sample essay / writing sample / writing model AND higher education. These databases were chosen because they are considered authoritative, ERIC in education and PsycINFO for educational psychology and psychology in general. We excluded a related term "worked example" from the keyword list because this construct involves facilitating complex problem-solving more than explicating assessment standards. No time range was set because we sought to identify as many relevant papers as possible.

We first identified 793 papers from the databases. Upon removing 28 duplicates, we read the abstracts of 765 papers to judge their relevancy under five inclusion criteria: (i) description of exemplar use procedure; (ii) empirical evidence on the educational effects of exemplars; (iii) a study in higher education; (iv) a peer-reviewed publication; (v) English as the language of the publication. During the screening, 574 papers failed to meet the inclusion criteria, leaving 191 for further assessment of their eligibility. We then excluded 147 papers after applying three exclusion criteria: (i) non-pedagogical use of exemplars (N=95); (ii) anecdotal report of teaching experience without research findings (N=45); (iii) no full text (N=7). We did not limit the type of publications during the search, but the identified book chapters, dissertations and conference papers either failed to meet the inclusion criteria or lacked full text. The first and second authors assessed the eligibility of the selected papers independently (inter-rater agreement 88%), with consensus reached through negotiation and involving the third author where necessary. Out of the remaining 44 publications, we excluded six due to weak research designs (e.g. the only use of a questionnaire with a small sample to investigate students' perceptions of exemplar use without data triangulation) (N=4) or unconvincing research arguments (e.g. claiming lack of improvement in students'

performance based on no changes in their ranking of scores) (N=2). To keep abreast of most recent publications during the drafting of manuscript, we conducted an additional search for relevant publications using the same set of inclusion and exclusion criteria, and two papers were added. We finally included 40 papers in the review. Due to the systematic nature of this review, we implemented the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) method. The PRISMA diagram in figure 1 summarises the literature search process and reports on the major decisions made in different stages.

Figure 1 PRISMA flowchart of literature search process



Data analysis

We extracted four categories of data from the selected studies: (i) research design; (ii) exemplar-related information; (iii) conditions of exemplar use; (iv) educational effects. The analysis was performed as follows. First, the research questions led us to identify three a priori themes (see the coding scheme in Table 1). Then, we defined categories to capture the essence of each theme when examining the studies, for example "rubrics use" and "engagement pattern" under "exemplar implementation". Codes were derived from the close reading of texts to label the defining features of each category. For instance, "beginning with exemplar analysis / grading" and "beginning with students' production of draft" were identified to illustrate "engagement pattern". A frequency count of each code was conducted to reveal its prevalence in the studies. The coding was followed by the writing of analytic memos to record the notable features of exemplar-based research designs, exemplar implementation and educational effects and to yield insights into productive exemplar-based research designs and pedagogy.

Table 1 Coding scheme

RQ	Theme	Category	Code
1	Research design	Research approach	Experimental; quasi-experimental; pre- & post-intervention evaluations; post-intervention evaluation; action research; teacher-research
		Data collection method	Survey; open-ended survey; focus groups; interviews; classroom observations; analysis of academic performance, assignment drafts, reflections or exemplars dialogue
2	Exemplar implementation	Rubrics use	Guide peer and self-assessment; support draft revision; co-construct criteria
		Engagement pattern	Beginning with exemplar grading/ analysis; beginning with students' production of draft

3	Educational effects	Perceptions	Useful; deeper understanding of standards and assessment process
		Academic performance	Improvement; equal performance; no improvement
		Self-regulated learning	Confidence; self-reflection; meta-cognitive awareness

The coding and analysis were initially led by the first author with the second author advising specifically on quantitative aspects of the research, and the third author providing input on the details of exemplars use. To enhance review rigour, a series of meetings between all authors were convened to raise issues, sharpen the analysis, clarify interpretations and reconcile differences of opinion. A diary was kept to document all the methodological decisions reached during the research process. Negotiation and discussion enabled us to resolve conflicting views, where necessary. The vast variations in research designs and exemplar-based practices impeded the performance of a meta-analysis and the calculation of effect sizes. We instead identified common patterns or themes from the selected studies, conducted frequency count and reported the results.

Results

This section begins with a description of the selected studies and then explores each research question in depth.

The included studies were mainly published after 2010 (N=36) (see Table 2 for details), showing a growing interest in exemplar-based interventions. Their sample sizes ranged widely from <50 to >1000, mostly conducted in the undergraduate context with 27 studies in the first two years of university education and 7 in the senior years. Written exemplars without annotated comments (N=30) were employed to facilitate the discussion of assessment standards. Written exemplars with annotated comments (N=8) were adopted under three conditions: (i) as self-explanatory material to illustrate standards (e.g. Scoles,

Huxham, and McArthur 2013); (ii) encouraging students to produce reflective insights during discussion (e.g. Hawe et al. 2020); (iii) consolidating students' learning of standards after discussion (e.g. Yucel et al. 2014). Generally two or three exemplars were used to illustrate a range of quality.

Table 2 Information about selected exemplar-based studies

Publication year	Before 2010	N=4
	2010 - 2013	N=11
	2014 - 2017	N=12
	2018 - 2021	N=13
Country	Australia	N=15
	UK	N=10
	US	N=5
	New Zealand	N=3
	Hong Kong	N=3
	China	N=2
	Israel	N=1
	Belgium	N=1
Academic level	Undergraduate (Years 1-2)	N=27
	Undergraduate (Years 3-5)	N=7
	Postgraduate	N=3
	Undergraduate & Postgraduate	N=2
	Associate Degree	N=1
Sample size	≤50	N=13
	51-100	N=8
	101-500	N=14
	501-1000	N=2
	>1000	N=3
Discipline	Science	N=9
	Education	N=8
	Social Sciences	N=8
	Business	N=6
	Language	N=5
	Law	N=2
	Philosophy	N=2
Type	Written (without annotations)	N=30
	Written (with annotations)	N=8
	Video	N=1
	Written and video	N=1

Number of exemplars	1	N=4
	2	N=9
	3	N=13
	4	N=4
	5	N=2
	6 or above	N=5
	Not specified	N=3
Quality	A range of quality	N=20
	Only excellent	N=11
	One strong and one weak	N=6
	Only borderline or weak	N=3

RQ1. What are the research designs of exemplar-based studies?

Six different research approaches were identified (see Table 3). There was a high number of post-intervention evaluation studies with two or three groups of participants receiving different exemplar conditions (N=12) and one-group post-intervention evaluation studies (N=15). Comparatively, there were low numbers of experimental (N=2) and quasi-experimental studies (N=1) as well as action or teacher research (N=6). Extracting conclusions from both types of post-intervention evaluation studies requires caution because of limitations in research designs.

Table 3 Research designs and types of data collection

Research approach	Post-intervention evaluation (one group)	N=15
	Post-intervention evaluation (two / three groups)	N=12
	Action research / teacher research	N=6
	Pre- & post-intervention evaluations	N=4
	Experimental	N=2
	Quasi-experimental	N=1
*Quantitative method	Analysis of academic performance	N=14
	Survey	N=13
*Qualitative method	Open-ended survey	N=13
	Focus groups / discussions	N=13
	Analysis of students' reflections	N=11
	Interviews	N=10
	Analysis of assignment drafts	N=5

	Analysis of exemplar dialogues	N=3
	Classroom observations	N=2

*The totality exceeds 40 because most studies employed more than one data collection method.

The post-intervention evaluation studies with two or three groups of participants usually compared the academic performance of the group(s) with exemplar use with that of the group without exemplar use. Importantly, there were six studies in which the exemplar and non-exemplar groups were from different cohorts (Handley and Williams 2011; Hendry and Anderson 2013; Wimshurst and Manning 2013; Yucel et al. 2014; Murphy 2015; Hill and West 2020). When both groups are not from the same cohort, the salient differences among the groups may influence the findings (Cohen et al. 2018). For example, Yucel et al. (2014) speculated that the better academic results in the non-exemplar group were due to the higher motivation and academic ability of the cohort and lecturers' marking variations between cohorts, illustrating the potential problem of non-equivalent group designs.

The one-group post-intervention evaluation studies usually investigated students' perceptions and their self-regulatory behaviour after exemplar use (e.g. Defeyter and McPartlin 2007; Bell, Mladenovic, and Price 2013; Bamber 2015; Aitken and Thompson 2018). Since there was no comparison with any other group and no pre-intervention evaluation, it was not feasible to determine whether students' positive perceptions and development of self-regulatory strategies were caused by exemplar use or other factors such as prior assessment experiences and academic capabilities (e.g. Bell et al., 2013; Bamber, 2015).

For data collection methods, all studies but one (Bacchus et al., 2020) assessed the effectiveness of exemplar-based practices with more than one method. This is positive as the convergence of multiple sources of evidence enables data triangulation and the revelation of different educational effects (Cohen et al., 2018). Surveys, open-ended surveys, interviews and focus groups were employed to explore students' perceptions and experiences of

exemplar use. The analysis of students' academic performance and their first and subsequent drafts could indicate if exemplar use led to changes in performance. The analysis of reflections could examine how students had assimilated assessment standards and had derived insights from exemplars.

RQ2. How are exemplar-based practices implemented?

Concerning exemplar implementation, the most salient characteristic was the concurrent use of rubrics and exemplars (N=25) to guide exemplar grading or analysis, scaffold peer and self-assessment of drafts and co-construct criteria with peers or teachers. For instance, in Rust et al.'s (2003) study, a rubric was given to participants to illustrate assessment standards so that they could apply the standards to grade exemplars. In Lipnevich et al.'s (2014) study, one of the experimental groups received a rubric and three exemplars to revise an assignment draft prior to submission. In Ayalon and Wilkie's (2020) study, participants completed an individual task, created task criteria for a rubric, refined the criteria through analysing multiple exemplars in pairs and reflected on the experience individually.

For implementation procedures, we identified two major engagement patterns: (i) engaging with exemplar grading or analysis; (ii) engaging with students' production of assignment draft (see Table 4 for details).

Table 4 Exemplar implementation procedures

Pattern 1: Engaging with exemplar grading / analysis (N=35), followed by:	1) Peer and / or teacher-student discussion of exemplar quality (N=17)
	2) Peer and / or teacher-student discussion of exemplar quality and peer and self-assessment of / teacher's feedback on assignment draft (N=6)
	3) Peer and / or self-assessment of draft with rubrics (N=5)
	4) Students' completion of draft and /or teacher's feedback (N=3)
	5) Students' raising questions about exemplars (N=2)
	6) Students' completion of a part of assignment and discussion of exemplar grades (N=1)
	7) Teacher's explanation of exemplar grades (N=1)

Pattern 2: Engaging with students' production of assignment draft (N=5), followed by:	1) Teacher's provision of a rubric and exemplars to scaffold draft revision (N=1)
	2) Students' comparison of draft with teacher's improved version of draft and an exemplar prior to revision (N=1)
	3) Exemplar analysis and students' construction and revision of criteria (N=1)
	4) Students' development of grading descriptors and comparison of their draft with exemplars (N=1)
	5) Exemplar marking, in-class discussion of the exemplar marks and co-development of draft improvement plan (N=1)

The first pattern engaged students with exemplar grading or analysis (N=35) in which they assigned a grade to exemplars or analysed the strengths and weaknesses of exemplars. The grading or analysis exercise was followed by different activities, with the aim of developing evaluative judgement, namely the capability to evaluate the quality of one's and peers' work (Tai et al. 2018). In 17 studies, the participants took part in peer and / or teacher-student discussion to explain their evaluative reasoning and mediate their understanding of assessment criteria through interaction (e.g. Hendry and Anderson 2013; To and Carless 2016; Tam 2021). To support their internalisation of criteria, six studies arranged peer- and / or self-assessment of drafts subsequent to the exemplar discussion (e.g. Lazar and Ellis 2011; Yucel et al. 2014) so that they could transfer evaluative skills to monitor their own performance. This engagement pattern emphasised students' development of evaluative judgement more than metacognitive monitoring during exemplar use.

The second pattern commenced with students' production of assignment drafts (N=5), followed by other exemplar-related arrangements to heighten their metacognitive awareness for self-monitoring. For example, in Yang and Zhang's (2010) study, participants compared their drafts with a teacher's improved version of the draft and an excellent exemplar and then discussed with peers the differences between their writing with the improved version and the exemplar. In Sambell and Graham's (2020) study, participants wrote grading descriptors for

teacher-derived criteria, compared their works with three exemplars and generated feedback for self-improvement. In Hill and West's (2020) study, participants compared their judgements of exemplars with the teacher's and co-developed a draft improvement plan. These studies shared commonalities in developing students' metacognitive awareness. The comparison of drafts with exemplars enabled students to notice their performance gap and generate internal feedback for self-monitoring.

RQ3. What are the educational effects of exemplar use?

This subsection discusses three major educational effects, including perceptions, changes in academic performance and SRL development.

Perceptions

Thirty studies documented students' perceived usefulness of exemplars, which makes it a frequently reported educational effect. The participants in these studies generally perceived exemplars to be effective in clarifying teachers' expectations of assessment tasks (e.g. Bamber 2015; Hawe et al. 2020), guiding task preparation (e.g. To and Carless 2016) and increasing their understanding of quality (e.g. Rust et al. 2003; Blair et al. 2014). They also valued exemplar discussion with teachers because it allowed them to seek teacher's guidance for assignment preparation (Tam 2021) and learn about self-evaluation skills (Hawe et al. 2019). Students' consistently positive perceptions of the usefulness of exemplars represents an argument for their increased use.

Academic performance

Fifteen studies evaluated students' changes in academic performance after exemplar use. Ten showed an improvement, four no changes and one decreased performance (See Table 5 for details). Ten studies reported an improvement when (i) students' marks or performance in the first draft / task was compared to those in the subsequent task (Yang and Zhang 2010; Jones et al. 2017; Broadbent et al. 2018); (ii) the results of the students with

exemplar support were compared to those of their counterparts without such support in the same cohort (Rust et al. 2003; Hendry et al. 2012; Scoles et al. 2013; Lipnevich et al. 2014); (iii) the performance of the cohort with exemplar use was compared to that of the previous cohort without exemplar use (Wimshurst and Manning 2013; Knight et al. 2019; Hill and West 2020). When probing into the factors for performance improvement, these studies engaged students actively in the process, for example self-assessment of draft after exemplar discussion (Hill and West 2020) or students' production of reflective insights from exemplar analysis (Jones et al. 2017). This suggests that students' active engagement with exemplars is useful in advancing students' performance.

Table 5 Students' academic performance after exemplar use

Study	Evaluation method	Academic performance
Rust et al. (2003)	Mark comparison within and between cohorts	Increase
Yang & Zhang (2010)	Performance analysis of first and revised drafts	
Hendry et al. (2012)	Mark comparison within cohort	
Scoles et al. (2013)	Mark comparison within cohort	
Wimshurst & Manning (2013)	Mark comparison between cohorts	
Lipnevich et al. (2014)	Mark comparison within cohort	
Jones et al. (2017)	Mark analysis of each task	
Broadbent et al. (2018)	Longitudinal analysis of marks	
Knight et al. (2019)	Mark comparison between cohorts	
Hill & West (2020)	Longitudinal analysis of grades; grade comparison between cohorts	
Handley & Williams (2011)	Mark comparison between cohorts	No change
Hendry & Anderson (2013)	Mark comparison between cohorts	
Smith et al. (2013)	Mark comparison within cohort	
Carter et al. (2019)	Mark comparison within cohort	
Yucel et al. (2014)	Mark comparison between cohorts	Decrease

Despite the evidence on performance improvement, the findings should be interpreted with caution for two reasons. First, since exemplars were utilised alongside rubrics, self- or peer assessment in four of the studies (Rust et al. 2003; Jones et al. 2017; Knight et al. 2019; Hill and West 2020), disentangling the combined effects was impossible. Most of the post-intervention evaluation studies compared the combined use of exemplars and other strategies in one group with another group without using exemplars and other strategies. Based on the findings, we could only conclude that the combined use of exemplars, rubrics, self- or peer-assessment appears to enhance students' performance but not that exemplar use directly leads to better performance. Second, we lacked conclusive evidence whether improvements were short-term or long-term. There were findings about improvements in subsequent assessment tasks in the same semester and one year after exemplar use (Rust et al. 2003; Broadbent et al. 2018). However, there were also findings about improvements in the subsequent task but not end-of-term examination (Wimshurst and Manning 2013), possibly because the genre differences between assignment and examination questions may have hindered students' transfer of exemplar insights.

Four studies showed no changes in performance after exemplar use (Handley and Williams 2011; Hendry and Anderson 2013; Smith et al. 2013; Carter et al. 2019). Except Smith et al. (2013), the remaining three studies compared students' performances with and without exemplar use only after an intervention. The lack of pre-intervention evaluation in these three studies may have affected the evaluation reliability. For the study with the decreased performance, Yucel et al. (2014) ascertained that students with peer review and exemplar use scored lower than their counterparts without exemplar use. This could be attributed to the contextual differences between cohorts, as mentioned in the section about research designs.

Self-regulated learning

Exemplar use was found to be useful in developing some subprocesses of SRL. For a systematic classification, we used the three phases of Zimmerman's (2000) SRL model to cast light on the specific subprocesses (see Table 6 for details).

Table 6 Development of self-regulated learning subprocesses after exemplar use

Study	Evaluation method	Phase: subprocess
Rust et al. (2003)	Interviews	Forethought: self-motivation beliefs
Dempsey et al. (2009)	Open-ended survey	Forethought: self-motivation beliefs
Yang & Zhang (2010)	Interviews	Self-reflection: self-judgement
Handley & Williams (2011)	Discussions	Self-reflection: self-judgement
Hendry et al. (2011)	Survey; focus groups	Forethought: task analysis; self-motivation beliefs
Hendry et al. (2012)	Survey; focus groups	Forethought: self-motivation beliefs
Hendry & Anderson (2013)	Survey; interviews	Forethought: self-motivation beliefs
Wimshurst & Manning (2013)	Reflections	Performance: self-observation
Hendry & Tomitsch (2014)	Survey; focus groups	Forethought: self-motivation beliefs
Murphy (2015)	Survey	Forethought: self-motivation beliefs
To & Carless (2016)	Open-ended survey; focus groups	Forethought: self-motivation beliefs
Hawe & Dixon (2017)	Interviews	Forethought: self-motivation beliefs Self-reflection: self-judgement
Jones et al. (2017)	Survey; reflections	Forethought: task analysis; self-motivation beliefs Performance: self-observation Self-reflection: self-judgement
Seery et al. (2017)	Surveys	Forethought: self-motivation beliefs
Hawe et al. (2019)	Survey; interviews; journal	Forethought: self-motivation beliefs Self-reflection: self-judgement
Hawe et al. (2020)	Interviews	Forethought: task analysis Self-reflection: self-judgement
Hill & West (2020)	Interviews; focus groups	Forethought: self-motivation beliefs Self-reflection: self-judgement
Sambell & Graham (2020)	Interviews	Self-reflection: self-judgement

Tam (2021)	Interviews; journals	Self-reflection: self-judgement

In the forethought phase (involving task analysis and self-motivation beliefs), a richer understanding of task requirements encouraged students to set realistic task goals and adopt appropriate task strategies. For task analysis (N=3), the participants in Jones et al.'s (2017) study reported using rubrics and exemplars to tailor their work to meet assessors' requirements. Those in Hawe et al.'s (2020) study referred to exemplars for structuring their assignment and determining the scope of literature to be included in the assignment. For self-motivation beliefs (N=13), participation in exemplar-based practices improved students' confidence in assignment preparation (e.g. Murphy 2015; Hill and West 2020) and raised their motivation to produce quality work (e.g. Hendry et al. 2012; Hendry and Tomitsch 2014).

In the performance phase (governing metacognitive monitoring of performance), a specific component of exemplar-based practices could facilitate students' self-observation (N=2). Making students write a self-reflective diary (Jones et al. 2017) and reflections (Wimshurst and Manning 2013) at the end of the practices aided students in synthesising and generating the insights from exemplar analysis and peer review for metacognitive monitoring.

In the self-reflection phase (supporting self-evaluation and management of task-related behaviour), for self-judgement (N=9), the comparison of students' drafts with exemplars helped students identify their strengths and weaknesses and prompted them to develop improvement plans. For example, in Yang and Zhang's (2010) study, the comparison of students' narrative with the teacher-improved version and a good exemplar enabled them to understand their problems and the aspects requiring improvement. In Sambell and Graham's (2020) study, the draft-exemplar comparison helped students generate internal feedback for self-regulation.

The above studies showed positive effects of exemplar use on SRL development, particularly in increasing students' self-motivation beliefs and enabling their self-judgement. However, we should be cautious in interpreting the findings because all of these studies used self-report data from surveys, interviews and focus groups to evaluate SRL development. Given the limitations of self-report data (see, for example, Fryer and Dinsmore 2020), we can only conclude that students hold a positive belief about the effect of exemplars on SRL development.

Discussion

This review examined the research designs of exemplar-based studies, exemplar implementation and the educational effects of exemplar use. The findings revealed frequent use of post-intervention evaluation studies and the integrated use of exemplars, rubrics, peer- or self-assessment. There was tentative evidence that exemplar use could improve performance. Students were positive about being exposed to exemplars and generally believed exemplars could enhance their understanding of quality and support their SRL development.

In examining the research designs and exemplar-based practices, we discovered that many studies incorporated exemplars, rubrics, peer or self-assessment into interventions. This carries implications for research designs. For quantitative studies, without controlling different variables it is impossible to disentangle the combined effects and provide robust evidence that it is specifically exemplar use rather than other strategies which benefits student learning. Plenty of these studies claimed to have found positive effects of exemplars, but sometimes this may overclaim more benefits of exemplar use than the data demonstrated. The field may benefit from additional true experimental (e.g. Lipnevich et al. 2014) and quasi-experimental studies (e.g. Smith et al. 2013), for example assigning exemplars only, rubrics only, peer or self-assessment only and the combination of all components to separate

experimental groups and compare participants' performance respectively. Future qualitative research could explore in more detail how students use exemplars and other strategies to inform their work and identify more productive learner strategies in the use of exemplars. Qualitative methods useful for pursuing such lines of further research could include self-reflective diaries (e.g. Jones et al. 2017) or think aloud protocols. Classroom observational research of exemplars implementation triangulated with interview data from participants would also be particularly useful (e.g. Smyth and Carless 2021).

In analysing the educational effects, the field lacks conclusive evidence that exemplars improve students' academic performance, although there was a tendency for positive effects (10 positive, 4 neutral and 1 negative). Inconsistent results may be explained by contextual factors or the choice of evaluation methods. The longitudinal analysis of students' scores across semesters (Broadbent et al. 2018; Hill and West 2020) and the comparison of students' performance or marks between first and final drafts (Yang and Zhang 2010; Lipnevich et al. 2014) represent potentially promising methods because the former enables researchers to trace students' transfer of exemplar insights to other tasks and the latter eliminates the possibility of genre differences between drafts. In contrast, although the across-cohort comparison of marks provides a larger data set, cohort differences in motivation and academic capability may have influenced the findings (Hendry and Anderson 2013; Yucel et al. 2014). For reliable evaluation of students' performance, we recommend researchers use longitudinal analysis of performance or analysis of drafts if possible. In the case of across-cohort comparison, it is advisable to adopt a pre-test post-test design to evaluate the variations between cohorts.

Our findings also revealed that exemplar use could scaffold students' SRL development, through raising their confidence in assignment preparation, metacognitive awareness and self-reflective capabilities. Benefits were identified in all three phases:

forethought (e.g. Hawe et al. 2020), performance (e.g. Jones et al. 2017) and self-reflection (e.g. Sambell and Graham 2020). This is an important finding because the main argument against exemplar use seems to be students' copying from exemplars (Handley and Williams 2011). The evidence that exemplars aid self-regulation appears to represent a significant counter-argument to that viewpoint.

The main pedagogical implication of our synthesis of exemplar implementation suggests that the usefulness of exemplars could be increased when students first produce assignment drafts prior to exposure to exemplars and then conduct peer review (Lazar and Ellis 2011; Yucel et al. 2014) or self-assessment (Rust et al. 2003; Hill and West 2020). This strategy supports students' internalisation of assessment criteria and academic self-regulation because evaluating peers' drafts and producing feedback require students to apply their understanding of criteria in making judgements. This evaluative experience can also be transferred to self-assessment. When carrying out self-assessment, students could compare their draft with good quality exemplars to identify performance gaps and raise metacognitive awareness (Yang and Zhang 2010; Sambell and Graham 2020). This substantiates the notion of reflective comparisons (Nicol 2021) which stresses students' production of internal feedback through comparing their draft with other works in the same genre. Significantly, the emphasis on internal feedback prompted by exemplar analysis might also be workload-efficient in reducing time-consuming or unproductive forms of teacher feedback (Carless 2020). Encouraging student ownership of the assignment task through producing drafts also carries potential to reduce the problems of inhibiting creativity and copying from exemplars.

This review has two main limitations. First, a small number of the selected studies presented low level of details in their methodology. This impeded analysing them with a higher level of certainty. Nevertheless, this paper offers useful insights into exemplar implementation and exemplar effects on different aspects of students' learning. Second, the

exemplar use in online learning was not sufficiently portrayed as the majority of the selected studies were conducted in a classroom setting. It would be illuminating for future researchers to explore exemplar implementation on a virtual platform to meet the needs in the post-pandemic period.

Conclusion

This review has found evidence that the integrated use of exemplars, rubrics, peer or self-assessment of drafts could enhance students' performance and SRL, though we need to be cautious due to the number of studies included and their use of evaluation methods. The integrated use has implications to researchers and practitioners. For quantitative researchers, there is a need for experimental or quasi-experimental designs to compare the effects of various components in exemplar-based practices. For qualitative researchers, it would be beneficial to use self-reflective diaries, classroom observations or other research instruments to examine how the combined use of exemplars and other strategies improves students' performance and self-regulated learning. For practitioners, the idea that students produce a draft before being exposed to exemplars represents a potentially important breakthrough. With advancements in research and pedagogical designs, exemplar use could assume a more prominent role in higher education pedagogy.

Declarations and ethics statements

No potential conflict of interest was reported by the authors.

Data availability statement

The data that support the findings of this study are available from the corresponding author, JT, upon reasonable request.

Funding

This research received no grant from any funding agency in public, commercial, or not-for-profit sectors.

References

The publications indicated with * were included as empirical evidence in this review.

*Aitken, A., and D. G. Thompson. 2018. “Using Software to Engage Design Students in Academic Writing.” *International Journal of Technology and Design Education* 28(3): 885-898. doi:10.1007/s10798-017-9413-4

*Ayalon, M., and K. J. Wilkie. 2020. “Developing Assessment Literacy through Approximations of Practice: Exploring Secondary Mathematics Pre-service Teachers Developing Criteria for a Rich Quadratics Task.” *Teaching and Teacher Education* 89. doi:10.1016/j.tate.2019.103011

*Bamber, M. 2015. “The Impact on Stakeholder Confidence of Increased Transparency in the Examination Assessment Process.” *Assessment and Evaluation in Higher Education* 40(4): 471-487. doi:10.1080/02602938.2014.921662

*Bacchus, R., E. Colvin, E. B. Knight, and L. Ritter. 2020. “When Rubrics aren’t Enough: Exploring Exemplars and Student Rubric Co-construction.” *Journal of Curriculum and Pedagogy* 17(1): 48-61. doi:10.1080/15505170.2019.1627617

*Bell, A., R. Mladenovic, and M. Price. 2013. “Students’ Perceptions of the Usefulness of Marking Guides, Grade Descriptors and Annotated Exemplars.” *Assessment and Evaluation in Higher Education* 38(7): 769-788. doi:10.1080/02602938.2012.714738

*Blair, A., A. Wyburn-Powell, M. Goodwin, and S. Shields. 2014. “Can Dialogue Help to Improve Feedback on Examinations?.” *Studies in Higher Education* 39(6): 1039-1054. doi:10.1080/03075079.2013.777404

*Bouwer, R., M. Lesterhuis, P. Bonne, and S. De Maeyer. 2018. “Applying Criteria to Examples or Learning by Comparison: Effects on Students' Evaluative Judgment and Performance in Writing.” *Frontiers in Education* 3. doi:10.3389/feduc.2018.00086

*Broadbent, J., E. Panadero, and D. Boud. 2018. “Implementing Summative Assessment with a Formative Flavour: A Case Study in a Large Class.” *Assessment and Evaluation in Higher Education* 43(2): 307-322. doi:10.1080/02602938.2017.1343455

Carless, D. 2020. “From Teacher Transmission of Information to Student Feedback Literacy: Activating the Learner Role in Feedback Processes”. *Active Learning in Higher Education*. doi:10.1177/1469787420945845

*Carless, D., and K. K. H. Chan. 2017. “Managing Dialogic Use of Exemplars.” *Assessment and Evaluation in Higher Education* 42(6): 930-941. doi:10.1080/02602938.2016.1211246

*Carter, R., E. Halcomb, L. M. Ramjan, N. J. Wilson, P. Glew, and Y. Salamonson. 2019. “Does the Use of Annotated Exemplars by Nursing Students Predict Academic Performance? A Cohort Study.” *Nurse Education Today 80*: 34-39. doi:10.1016/j.nedt.2019.06.003

Cohen, L., L. Manion, and K. Morrison. 2018. *Research Methods in Education* 8th ed. London: Routledge.

*Defeyter, M. A., and P. L. McPartlin. 2007. “Helping Students Understand Essay Marking Criteria and Feedback.” *Psychology Teaching Review* 13(1): 23-33. http://www.bps.org.uk/ptr

*Dempsey, M. S., L. M. PytlikZillig, and R. H. Bruning. 2009. “Helping Preservice Teachers Learn to Assess Writing: Practice and Feedback in a Web-based Environment.” *Assessing Writing* 14(1): 38-61. doi:10.1016/j.asw.2008.12.003

Fryer, L. K., and D. Dinsmore. 2020. “The Promise and Pitfalls of Self-report: Development, Research Design and Analysis Issues and Multiple Methods.” *Frontline Learning Research* 8(3): 1-9. doi:10.14786/flr.v8i3.623

Garg, A. X., D. Hackam, and M. Tonelli. 2008. “Systematic Review and Meta-analysis: When One Study is Just Not Enough.” *Clinical Journal of the American Society of Nephrology* 3(1): 253-260. Doi:10.2215/CJN.01430307

*Handley, K., and L. Williams. 2011. “From Copying to Learning: Using Exemplars to Engage Students with Assessment Criteria and Feedback.” *Assessment and Evaluation in Higher Education* 36(1): 95–108. doi:10.1080/02602930903201669

*Hawe, E., and H. Dixon. 2017. “Assessment for Learning: A Catalyst for Student Self-regulation.” *Assessment and Evaluation in Higher Education* 42(8): 1181-1192. doi:10.1080/02602938.2016.1236360

*Hawe, E., U. Lightfoot, and H. Dixon. 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. doi:10.1080/0309877X.2017.1349894

*Hawe, E., H. Dixon, J. Murray, and S. Chandler. 2020. “Using Rubrics and Exemplars to Develop Students’ Evaluative and Productive Knowledge and Skill.” *Journal of Further and Higher Education*. doi:10.1080/0309877X.2020.1851358

*Hendry, G. and J. Anderson. 2013. “Helping Students Understand the Standards of Work Expected in an Essay: Using Exemplars in Mathematics Pre-service Education Classes.” *Assessment and Evaluation in Higher Education* 38(6): 754-768. doi:10.1080/02602938.2012.703998

*Hendry, G., N. Bromberger, and S. Armstrong. 2011. “Constructive Guidance and Feedback for Learning: The Usefulness of Exemplars, Marking Sheets and Different Types of Feedback in a First Year Law Subject.” *Assessment and Evaluation in Higher Education* 36(1): 1-11. doi:10.1080/02602930903128904

*Hendry, G., S. Armstrong, and N. Bromberger. 2012. “Implementing Standards-based Assessment Effectively: Incorporating Discussion of Exemplars into Classroom

Teaching." *Assessment and Evaluation in Higher Education* 37(2): 149–161. doi:10.1080/02602938.2010.515014

*Hendry, G., and M. Tomitsch. 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. doi:10.1007/s10798-013-9256-6

*Hendry, G., P. White, and C. Herbert. 2016. "Providing Exemplar-based 'Feedforward' Before an Assessment: The Role of Teacher Explanation." *Active Learning in Higher Education* 17(2): 99-109. doi.:10.1177/1469787416637479

*Hill, J., and H. West. 2020. "Improving the Student Learning Experience through Dialogic Feed-forward Assessment." *Assessment and Evaluation in Higher Education* 45(1): 82-97. doi:10.1080/02602938.2019.1608908

*Jones, L., B. Allen, P. Dunn, and L. Brooker. 2017. "Demystifying the Rubric: A Five-step Pedagogy to Improve Student Understanding and Utilisation of Marking Criteria." *Higher Education Research and Development* 36(1): 129-142. doi:10.1080/07294360.2016.1177000

*Knight, S., A. Leigh, Y. C. Davila, L. J. Martin, and D. W. Krix. 2019. "Calibrating Assessment Literacy through Benchmarking Tasks." *Assessment and Evaluation in Higher Education* 44(8): 1121-1132. doi:10.1080/02602938.2019.1570483

*Lazar, G., and E. Ellis. 2011. "Genre as Implicit Methodology in a Collaborative Writing Initiative." *International Journal of English Studies* 11(1): 155-168. doi:10.6018/ijes/2011/1/137151

*Lipnevich, A. A., L. N. McCallen, K. P. Miles, and J. K. Smith. 2014. "Mind the gap! Students' Use of Exemplars and Detailed Rubrics as Formative Assessment." *Instructional Science* 42(4): 539-559. doi:10.1007/s11251-013-9299-9

*Macbeth, K. P. 2010. “Deliberate False Provisions: The Use and Usefulness of Models in Learning Academic Writing.” *Journal of Second Language Writing* 19(1): 33-48. doi:10.1016/j/jslw.2009.08.002

Mallett, R., J. Hagen-Zanker, R. Slater, and M. Duvendack. 2012. “The Benefits and Challenges of Using Systematic Reviews in International Development Research.” *Journal of Development Effectiveness* 4(3): 445-455. doi:10.1080/19439342.2012.711342

*Murphy, C. 2015. “The Use of Peer Modeling to Increase Self-efficacy in Research Methods Courses.” *Journal of Political Science Education* 11(1): 78-93. doi:10.1080/15512169.2014.985107

Nicol, D. 2021. “The Power of Internal Feedback: Exploiting Natural Comparison Processes.” *Assessment and Evaluation in Higher Education* 46(5): 756-778. doi:10.1080/02602938.2020.1823314

O’Donovan, B., M. Price, and C. Rust. 2008. “Developing Student Understanding of Assessment Standards: A Nested Hierarchy of Approaches.” *Teaching in Higher Education* 13(2): 205–217. doi:10.1080/13562510801923344

*Orsmond, P., S. Merry, and K. Reiling. 2002. “The Use of Exemplars and Formative Feedback When Using Student Derived Marking Criteria in Peer and Self-assessment.” *Assessment and Evaluation in Higher Education* 27(4): 309-323. doi:10.1080/0260293022000001337

Panadero, E., and A. Jonsson. 2020. “A Critical Review of the Arguments Against the Use of Rubrics.” *Educational Research Review* 30. doi:10.1016/j.edurev.2020.100329

*Rust, C., M. Price, and B. O’Donovan. 2003. “Improving Students’ Learning by Developing Their Understanding of Assessment Criteria and Processes.” *Assessment and Evaluation in Higher Education* 28(2): 147–164. doi:10.1080/02602930301671

Sadler, R. 2005. “Interpretations of Criteria-based Assessment and Grading in Higher Education.” *Assessment and Evaluation in Higher Education* 30(2): 175-194. doi:10.1080/0260293042000264262

Sadler, R. 2009. “Indeterminacy in the Use of Preset Criteria for Assessment and Grading.” *Assessment and Evaluation in Higher Education* 34(2): 159-179. doi:10.1080/02602930801956059

*Sambell, K., and L. Graham. 2020. ““We Need to Change What We’re Doing.” Using Pedagogic Action Research to Improve Teacher Management of Exemplars.” *Practitioner Research in Higher Education Journal* 13(2): 3-17. https://ojs.cumbria.ac.uk/index.php/prhe/article/view/631

*Scoles, J., M. Huxham, and J. McArthur. 2013. “No Longer Exempt from Good Practice: Using Exemplars to Close the Feedback Gap for Exams.” *Assessment and Evaluation in Higher Education* 38(6): 631-645. doi:10.1080/02602938.2012.674485

*Seery, M. K., H. Y. Agustian, E. D. Doidge, M. M. Kucharski, H. M. O’Connor, and A. Price. 2017. “Developing Laboratory Skills by Incorporating Peer-review and Digital Badges.” *Chemistry Education Research and Practice* 18(3): 403-419. doi:10.1039/C7RP00003K

*Smith, C. D., K. Worsfold, L. Davies, R. Fisher, and R. McPhail. 2013. “Assessment Literacy and Student Learning: The Case for Explicitly Developing Students ‘Assessment Literacy’.” *Assessment and Evaluation in Higher Education* 38(1): 44-60. doi:10.1080/02602938.2011.598636

Smyth, P., and D. Carless. 2021. “Theorising how Teachers Manage the Use of Exemplars: Towards Mediated Learning from Exemplars”. *Assessment and Evaluation in Higher Education* 46(3): 393-406. doi:10.1080/02602938.2020.1781785

Tai, J., R. Ajjawi, D. Boud, P. Dawson, and E. Panadero. 2018. “Developing evaluative judgement: Enabling students to make decisions about the quality of work.” *Higher Education* 76(3): 467-481. doi:10.1007/s10734-017-0220-3

*Tam, A. C. F. 2021. “Undergraduate Students’ Perceptions of and Responses to Exemplar-based Dialogic Feedback.” *Assessment and Evaluation in Higher Education* 46(2): 269-285. doi:10.1080/02602938.2020.1772957

*To, J., and D. Carless. 2016. “Making Productive Use of Exemplars: Peer Discussion and Teacher Guidance for Positive Transfer of Strategies.” *Journal of Further and Higher Education* 40(6): 746–764. doi:10.1080/0309877X.2015.1014317

To, J., and Y. Liu. 2018. “Using Peer and Teacher-student Exemplar Dialogues to Unpack Assessment Standards: Challenges and Possibilities.” *Assessment and Evaluation in Higher Education* 43(3): 448-460. doi:10.1080/02602938.2017.1356907

Torrance, H. 2007. “Assessment as Learning? How the Use of Explicit Learning Objectives, Assessment Criteria and Feedback in Post-secondary Education and Training can Come to Dominate Learning.” *Assessment in Education* 14(3): 281-294. doi:10.1080/09695940701591867

*Wimshurst, K., and M. Manning. 2013. “Feed-forward Assessment, Exemplars and Peer Marking: Evidence of Efficacy.” *Assessment and Evaluation in Higher Education* 38(4): 451-465. doi:10.1080/02602938.2011.646236

*Yang, L., and L. Zhang. 2010. “Exploring the Role of Reformulations and a Model Text in EFL Students’ Writing Performance.” *Language Teaching Research* 14(4): 464-484. doi:10.1177/1362168810375369

*Yucel, R., F. L. Bird, J. Young, and T. Blanksby. 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 and Evaluation in Higher Education* 39(8): 971-986. doi:10.1080/02602938.2014.880400

*Wu, Z. 2019. "Understanding Students' Mimicry, Emulation and Imitation of Genre Exemplars: An Exploratory Study." *English for Specific Purposes 54*: 127-138. doi:10.1016/j.esp.2019.02.002

Zimmerman, B. J. 2000. "Attaining Self-regulation: A Social Cognitive Perspective." In *Handbook of Self-regulation,* edited by M. Boekaerts, P. R. Pintrich, and M. Zeidner, 13-39. San Diego: Academic Press.