European Journal of Engineering Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ceee20

Mapping out the landscape of literature on assessment in engineering education

Jonathan G. Halls, Carmen Tomás, John S. Owen & Kamel Hawwash

To cite this article: Jonathan G. Halls, Carmen Tomás, John S. Owen & Kamel Hawwash (2022) Mapping out the landscape of literature on assessment in engineering education, European Journal of Engineering Education, 47:3, 373-393, DOI: 10.1080/03043797.2021.2009775

To link to this article: https://doi.org/10.1080/03043797.2021.2009775

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 29 Nov 2021.

Submit your article to this journal

Article views: 599

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ceee20

EUROPEAN JOURNAL OF ENGINEERING EDUCATION
2022, VOL. 47, NO. 3, 373–393
https://doi.org/10.1080/03043797.2021.2009775

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Mapping out the landscape of literature on assessment in engineering education

Jonathan G. Halls [a], Carmen Tomás [a,b], John S. Owen [c] and Kamel Hawwash [d]

[a]School of Education, University of Nottingham, Nottingham, UK; [b]Educational Excellence Team, University of Nottingham, Nottingham, UK; [c]Faculty of Engineering, Department of Civil Engineering, University of Nottingham, Nottingham, UK; [d]School of Civil Engineering, University of Birmingham, Birmingham, UK

ABSTRACT

Over recent decades, literature on assessment in higher education has intensified generating a wealth of frameworks to inform practice. Generic frameworks for assessment practice are sometimes perceived as missing subject-specific considerations. This literature review proposes to (a) map the current landscape of assessment in engineering education and to (b) help drive the field forwards by identifying elements of assessment that require discipline-specific consideration as a foundation to formulate good practice. Sources were identified using a broad set of keywords related to assessment in engineering education. Inclusion criteria considered papers about university-level education and were published, in English, between 2012 and 2018. The review establishes that much literature has focused on design, accreditation and marking with much less literature on key concerns in practice such as workplace assessment, student engagement and programme level design. Based on the results, recommendations are made for research areas where greater focus is needed to advance further engineering specific insights.

ARTICLE HISTORY
Received 16 February 2021
Accepted 15 November 2021

KEYWORDS
Assessment; course assessment; engineering education; literature review; systematic review

1. Introduction

Quality of teaching and assessment is becoming ever more important in higher education institutions, as demonstrated, in the UK, by heightened scrutiny in the regulatory environment (i.e. the Teaching Excellent Framework) and recent calls, in Europe, to improve student-centred learning (e.g. Todorovski, Nordal, and Isoski 2015). For staff working on engineering programmes of study, the growing literature on engineering education provides guidance and evidence-based practice to support their teaching and assessment activities. Even focusing solely on assessment, there has been an increased focus on assessment-related publications in the last few decades (Nawaz and Strobel 2016). However, with articles being published at an increasing rate, it can be difficult to develop a clear overview of a field of study. We propose to synthesise current literature trends and findings to better inform practice and identify new directions. Note that the term *literature* is used in this paper to refer to both (a) empirical articles evaluating new and existing assessment methods and (b) articles/case studies that describe assessment methods used in a certain situation.

CONTACT Jonathan G. Halls jonathan.halls2@nottingham.ac.uk
Supplemental data for this article can be accessed https://doi.org/10.1080/03043797.2021.2009775.

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

1.1 Background

The past two decades have seen a growth of subject-general literature on assessment as a result of the growing interest in both practice and research. It is subject-general in that it applies to all subject areas (e.g. engineering education) as the literature and systematic reviews converge on sending key messages, rather than prescribing subject-specific activities or fixed design guidelines. For example, the Higher Education Academy (HEA) framework, *Transforming assessment in higher education* (HEA 2018) outlines six tenets: promoting assessment for learning (formative assessment), developing assessment that is fit for purpose, recognising that assessments lack precision, constructing standards in communities, integrating assessment literacy into course design and ensuring professional judgements are reliable. These are underpinned by three areas of focus: innovative assessment, feedback practices, and self- and peer-assessment. The framework details the importance of giving priority to programme- and institutional-level review and assessment design to enhance assessment practices.

Another example is the frequently referenced Evans Assessment Tool (Evans 2016). This toolkit outlines three core dimensions: assessment literacy, assessment feedback and assessment design, each with four sub-dimensions. These dimensions are interconnected and should be considered together (i.e. programme level review). There exist other frameworks with similar messages and breadth of scope. Additional frameworks cover a certain sub-area of assessment practice. For example, the Assessment of Learning Outcomes in Higher Education framework (Tremblay, Lalancette, and Roseveare 2012) considers aspects of design and in particular, alignment of learning outcomes. Other frameworks and reviews offer a broader perspective discussing the fundamentals of assessment and feedback in a comprehensive manner (Jackel, Pearce, and Radloff 2017).

Furthermore, assessment, and associated literature, is complex. General higher education literature reviews already establish that best practice amounts to the systematic consideration of practice on a range of elements overall. These include work on aspects of design covering various elements: assessment design (Bloxham and Boyd 2007); credit weightings (Galvin, Noonan, and O'Neill 2012); constructive alignment (Biggs and Tang 2011); design of MCQs (Brame 2013) and design of marking rubrics (Brookhart 2018; Suskie 2018). More recently, there has been a growing focus on programme level assessment design strategies (Jessop and Tomás 2017; Hartley and Whitfield 2012) and programme level reviews (TESTA 2019; Tomás and Jessop 2018). Programme level considerations are recognised as key to overcoming many of the design challenges in practice associated with modularisation (assessment overloads, inconsistent practices).

Other frameworks focus more specifically on student engagement, focussing on feedback. For example, a range of frameworks are available: developing evaluative judgment (Tai et al. 2018; Boud et al. 2018); feedback (Evans 2016; Nicol 2010) and engagement with feedback (Winstone et al. 2017). Additional literature reviews have provided insights into a range of related elements: self-, peer- and co- assessment (Dochy, Segers, and Sluijsmans 1999); formative vs summative assessment (Lau 2015); and the use of assessment rubrics in formative assessment (Panadero and Jönsson 2013), to improve validity (Reddy and Andrade 2010), develop students' self-efficacy (Panadero and Romero 2014) and enhance feedback (Cockett and Jackson 2018).

While these systematic and selective literature reviews have been conducted for a range of educational practices, to both summarise findings and outline frameworks for design, a search of the literature returned no such synthesis of current work on assessment in engineering education. It is important to identify what is currently happening in engineering education. More importantly, the assessment literature and multiple frameworks, converge on reflecting the importance of assessment as a process. Consideration of multiple elements in the assessment life-cycle is what ultimately is considered good practice. We propose to examine whether the engineering literature reflects the main elements of existing assessment frameworks from generic higher education literature on assessment. If this relationship is not reflected in the literature, it could indicate that engineering education typically draws on research conducted in engineering education settings, rather than other academic fields (e.g. psychology or medical education) or generic frameworks. Such practices could limit the development of assessment practices in engineering, as important assessment principles from other areas are not incorporated into engineering education. This would be concerning, as the importance of enhancing assessment in higher education has been increasing in recent years.

1.2 Focus of review

The focus of the review is to summarise the topics and methods used in the current literature on assessment in engineering education. It is not a systematic review that examines the key findings of all papers reviewed. Key theoretical approaches to assessment practices are set in general higher education reviews which are referred to in our review. Rather, the aim is to summarise current trends in engineering education literature on assessment. Our aim is to help direct research efforts into assessment practices in engineering to enable building an evidence base to formulate its good practice. A foundation for engineering to build its own evidence frameworks for good practice is to ensure research on assessment practices address all foundational elements of assessment practice known from the generic literature.

1.3 The assessment life-cycle stages and key elements: a framework for our review

Drawing from the existing literature reviews quoted above, we created a framework for this review. Many of the existing frameworks formulate principles or synthesise insights on good practice. Our starting point considered comprehensive frameworks (e.g. Evans 2016; Suskie 2018) which demonstrate that good practice, overall, consists of a range of considerations that focus on multiple elements: design of individual modes and at the programme level, engaging students in assessment, making judgements (marking and moderation) and review processes. Assessment practice is complex and multidimensional due to the many functions it fulfils (Bloxham 2014), including driving student learning, certifying student learning and qualify assurance of standards of learning. This multidimensional set of considerations has led to the formulation of assessment life-cycles in many instances to help clarify good practice (Price et al. 2012; Bloxham, Hughes, and Adie 2016; Bearman et al. 2016; Boud et al. 2018; Tomás and Scudamore 2014; HEA 2012; MMU 2019).

The framework used in this review is detailed in Table 1. The framework draws on existing assessment life-cycles from existing literature reviews, to explore five key building blocks of assessment life-cycles that consider key stages (see Table 1, Level 1): (1) design at the instrument level; (2) design at the programme level; (3) communication (incl. both support and engagement of students); (4) marking (and moderation); (5) review (of task, rubrics and results). These five key blocks are further divided into specific sub-areas (see Table 1, Level 2), termed assessment life-cycle elements or, simply, assessment elements. These assessment elements reflect key areas of consideration for practitioners in the literature that reflect practice. In the interest of having a manageable framework, we synthesised common areas as broader elements for the purpose of this review. The sources referenced above discuss good practice and give specific guidelines. We have focused on elements as a starting point to indicate a broad range of fundamental considerations for good practice identified in contemporary literature. Assessment elements, therefore, provide a foundation to examine and review the evidence base to date in assessment practices in engineering.

In addition to the elements identified in the generic literature base, discussions were held with colleagues, across two different institutions, to identify important aspects of assessment for those working in engineering education. For example, project-based assessment, workplace-based assessment and processes related to accreditation are common areas of literature in engineering education. Therefore, these elements of assessment were incorporated into the coding scheme (Table 1) as part of the broader category of instrument level design (block 1). This was considered important as the first attempt to lay a foundation for an engineering-specific assessment framework that engineering educators would see as relevant. These do not link explicitly to existing frameworks but, for example, Suskie (2018) also provides considerations for the design of specific assessment instruments.

Table 1. Coding scheme for assessment element.

Level 1	Level 2	Code	Primary article focus
Instrument level design	University-based	AE 1	Design of assessment used in university settings, e.g. Exams, Essays, Labs, etc.
	Workplace-based	AE 2	Design of assessment used on placement.
	Professional skills	AE 3	Design of assessment of professional skills
	Formative assessment	AE 4	Design of assessment for developing student assess. lit.
	e-design	AE 5	Assessments using digital tools that result in non-traditional production/ outcomes
Programme level design	Constructive alignment	AE 6	The linking of module LOs, assess. design and instruction.
	Evaluative criteria	AE 7	Design of/relationship between assessment criteria and design
	Quantitative load	AE 8	Number of assessments across year/programme
	Chronologic load	AE 9	Timing of assessment, incl. formative
	Qualitative load	AE 10	How assessments cover LOs across programme, incl. accreditation.
	Performance standards	AE 11	Development of performance standards across a programme
	Administrative	AE 12	Procedures regarding submissions, returns and record-keeping (incl. digital)
	Assessment culture	AE 13	Staff assessment literacy, incl. strategies/policy
Communication (support and engagement)	Communication about assessment	AE 14	How students are informed about assessment procedures, criteria, timelines, etc.
	Student understanding of assessment	AE 15	Staff-led methods for developing students assess. literacy (e.g. workshops, tutorials).
Marking	Pre-marker training	AE 16	Staff shared-understanding of marking, incl. application of rubric
	Marking	AE 17	Marking processes
	Feedback construction	AE 18	The construction and volume of feedback given to students
	Post-marking moderation	AE 19	Inter and intra marker consistency
Review (of task, rubrics and results)	Inter and intra student validity	AE 20	The external validity of marks (e.g. how well marks relate to other tasks) and consistency of marks with other performances (for an individual)
	Rationale for assessment task	AE 21	The evidence base for task analysis and structural validity (post-hoc analyses)

The second key block in Table 1, specifies elements in relation to programme level design as this is recognised as key to improving assessment practice. The EAT framework (Evans 2016), introduces fourteen principles for practice, based on a review of the literature comprising elements of design at programme level and student engagement. The novelty of this framework is a set of programme level considerations, although Suskie (2018) has suggested similar considerations. These principles have been converted into identifiable elements in the assessment life-cycle. Many of Evans' (and Suskie's) suggestions for good practice are expressed as detailed checklists to practitioners. They overlap with known elements such as constructively aligning, setting criteria and developing standards (Biggs and Tang 2011). However, we have added additional elements that highlight programme level considerations.

Drawing on recent discussions about assessment load in programmes of study (TESTA 2019; Tomás and Jessop 2018) we added three key considerations of student assessment load at programme level: quantitative, qualitative, and chronologic. These are considerations in practice deemed to put students at the centre. Considering the total assessment load for a student (quantitative), the mapping of assessment types to outcomes (qualitative) and, lastly, the frequency of assessments and the bunching of assessment deadlines at programme level (chronologic) are all measures deemed important to improve student learning. Finally, administrative and cultural aspects were added under programme level design, as these are also key elements in two of the most comprehensive guides (Evans 2016; Suskie 2018). Administrative and cultural aspects relate to consistency in submissions (e.g. tools used), along with training and support policies.

The third key block in Table 1, communication, is divided into two main elements. One relates to communication about assessment and the other to 'engagement' in activities promoting student assessment literacy (student understanding of assessment). These two broad elements encompass multiple considerations from Evans (2016) and Suskie (2018), along with aspects from more focused reviews on developing student autonomy and self-regulated learning (e.g. Boud et al. 2018; Dochy, Segers, and Sluijsmans 1999; Tai et al. 2018; Panadero and Jönsson 2013). These two elements synthesise a whole range of considerations in practice. Readers can follow up on the literature suggested for detailed discussions and guidelines. For simplicity, the elements included consider whether students receive information or whether they are engaged actively.

The fourth key block, in Table 1, broadly focuses on marking. The elements within this block reflect an ideal set of considerations for practice drawn from the extensive literature on how to increase the validity and consistency of marking. Ensuring familiarisation, intra- and inter-marker reliability are all key recommendations for good practice (Brown 2001; Biggs and Tang 2011; Messick 1994, 1996). Pre-marking training is a widely recommended technique to improve consistency in the understanding and application of criteria. Post-marking moderation checks are well established in practice and they serve to assert adequate application of criteria and standards. Different sources may recommend different approaches, yet, most guides recommend taking these elements into consideration for good practice.

The fifth block in Table 1 covers review of assessments. What constitutes review will vary by assessment type, however, review plays a central role as part of a culture of routine improvement of assessment designs and practice (Messick 1996; Tavakol and Dennick 2017; Suskie 2018; van der Vleuten et al. 2015). While it is acknowledged that many different approaches may be adopted, included in the framework are two common elements at this life-cycle stage. The first element focuses on checking consistency of performance across tasks, raters and occasions. This is a common recommendation in the literature base (Baartman et al. 2007; Dochy 2009). The second element involves reviewing the assessment task.

2. Methods

There exist many methods for conducting a literature review. The most robust and comprehensive is the systematic review (see Davies 2000; Glass 2016). Typically, systematic reviews provide a synthesis of all literature reviewed and compare study findings. As this paper only aims to provide an overview of the field it, therefore, is not termed a systematic review. A true systematic review of the literature, detailing and comparing all study findings, would have been problematic, as the existing literature explored a wide range of topics that rarely overlapped.

While the output is not termed a systematic review, the methods used draw on systematic review methods. Such reviews follow a fixed research design with data collection methods stipulated in advance. The work here is based on guidance by Gough (2007). Detailed here are key decisions made surrounding inclusion criteria, identification of potential sources, study selection, the data extraction process and data analysis methods. As is common practice, reporting methods is comprehensive to ensure transparency and replicability.

2.1 Inclusion and exclusion criteria

The scope of what could be included in a review for the research aim is broad. However, the main criterion for inclusion was that the literature must cover assessment in relation to engineering education. A systematic methodology was chosen to select sources for inclusion, to qualify sources had to include all three of the following:

(1) The term *engineering*;
(2) The term *education*;

(3) Any of the following assessment-related terms: *assessment, assignment, accreditation, capstone, concept inventory, dissertation, evaluation, exam, examination, grading, homework, multiple choice, placement, portfolio, project, report* or *test*.

The keywords in point 3 were generated from recent abstracts of articles related to assessment in higher education, sourced from recent editions (2013–2018) of seven engineering education journals (Advances in Engineering Education, Engineering Studies, European Journal of Engineering Education, International Journal of Engineering Education, International Journal of Engineering Pedagogy, Journal of Engineering Education, Journal of Professional Issues in Engineering Education and Practice), along with additional search terms from the recent *taxonomy of engineering education keywords* (Finelli, Borrego, and Rasoulifar 2016).

The review focused on the period 2012–2018. This cut-off point was chosen based on (a) the end of the research project (2018) and (b) the publication of the HEA report *A Marked Improvement* (HEA 2012). Published in 2012, this influential report argued for sweeping changes to assessment across the higher education sector. Therefore, it is proposed to represent the beginning of a sea change in assessment practices, in the UK. We acknowledge that the rationale for cut-off point has a UK bias; however, the findings and scope of articles reviewed should apply to an international context. The aim here was to capture the current state of the field. It was assumed that principles of good practice detailed in pre-2012 literature that had stood the test of time would be present in the post-2012 literature.

In addition to (a) search terms and (b) publication date, there were three exclusion criteria:

- Irrelevant population, i.e. primary school children;
- Irrelevant source format, for example, letters to the editor;
- Not published in English.

2.2 Identification of sources

A range of potential sources were identified to provide information about best practice in engineering education. These sources were those that were expected to be returned during study selection, through the search strategy outlined below. Therefore, the creation of these lists was a preventative measure to avoid overlooking potential sources during data collection.

These sources were identified by exploration of *The SCImago Journal & Country Rank* research-webtool (https://www.scimagojr.com/) and through discussion with colleagues working in engineering education. Also reviewed were several high impact journals that publish work on assessment in higher education, as these journals may have contained articles specifically covering assessment in engineering education. These sources were identified by exploration of *The SCImago Journal & Country Rank* and through discussion with colleagues working in assessment enhancement in higher education (see Supplemental data for full list of journals).

2.3 Study selection (search strategy)

The search strategy was twofold. During Step 1 a range of research databases was searched, using the predetermined search terms (see Section 2.2). Databases included:

- EBSCO – ERIC (an education research database);
- Web of Knowledge (a large research database);
- The EPPI Centre (a systematic review database).

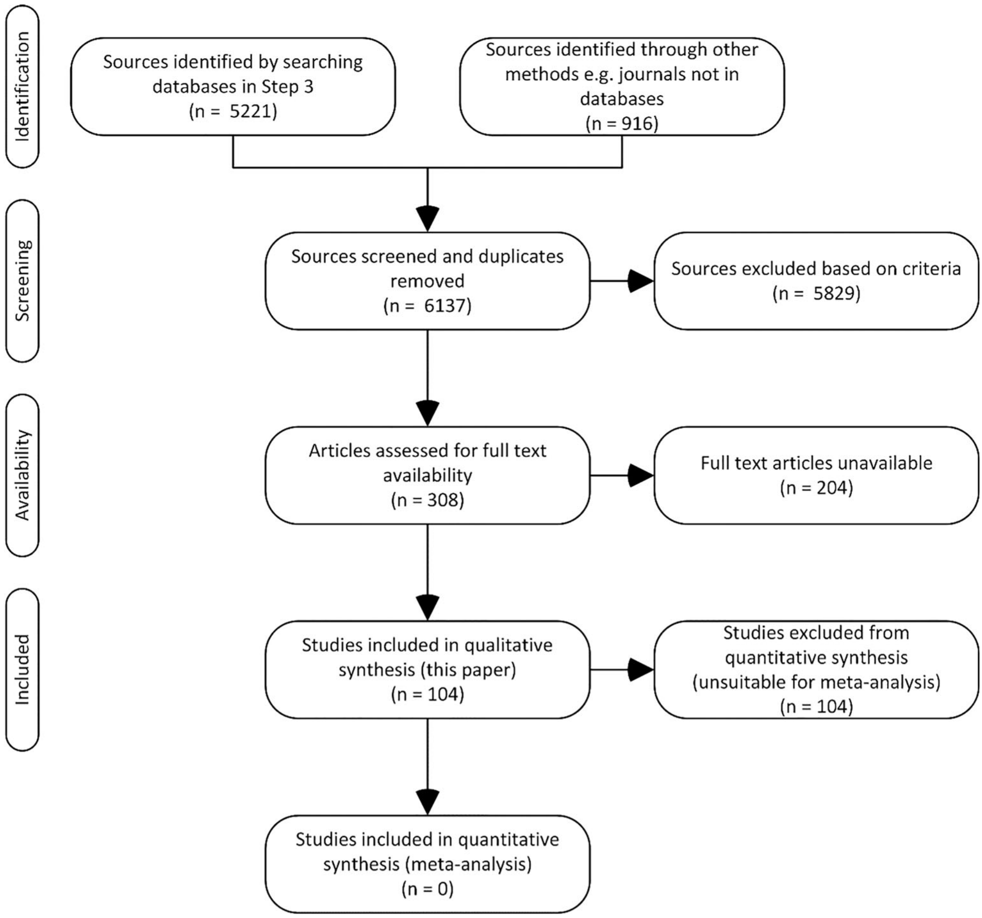

Figure 1. Flow chart for inclusion of sources: outlined are the number of sources returned during the initial database and journal searches, and the number of sources excluded at three different data review points (screening based on exclusion criteria, article availability and possible inclusion in meta-analysis).

During Step 2 if any of the potential sources (see Section 2.2) were not hosted by at least one of the databases then that journal's website was searched.

For all returned sources, the article abstracts were reviewed to ascertain if they met the inclusion and exclusion criteria. The broad search terms used resulted in a large number of articles being initially included, however, review of the abstracts revealed the majority of articles were not about assessment in engineering education. For example, a considerable proportion of the returned sources were articles that focused on assessing the viability of specific engineering projects in industry. Provided that they met the criteria, articles were added to a database. If the full-text source was not available through institutional access or open-access online repositories, article authors were contacted and asked to provide the full-text. However, due to time constraints, there was a 3-month cut-off point for inclusion of such articles. It was not within budget to pay for access to individual articles. Figure 1 summarises the results of the search strategy (for the full list of included articles see Supplemental data).

2.4 Data extraction

The 104 studies that met all inclusion criteria, and were available, were read and coded for a range of variables. Supplemental data provides a summary of variables not discussed in the paper (e.g. engineering discipline). Variables covered here include:

1. *Author type*, the professions of article authors: engineering educators, non-engineering educators and a mixed team.
2. *Journal type*, whether the journal in which the article was published was engineering education specific.
3. *Article references*, the number of engineering-specific and non-engineering specific references in an article, based on whether the reference made explicit reference to 'engineering' or 'engineers'.
4. *Assessment life-cycle element,* detailed below.
5. *Article method*, detailed below.

Variables 1–3 were included to explore the link between theory and practice in schools of engineering education. That is, to investigate how practitioners (e.g. lectures in engineering education) engage with the wider education literature base of which engineering education is a part. The variables of author type, journal type and article references serve as proxies for establishing the level of engagement with wider education literature. Higher levels of mixed author teams, references to non-engineering specific literature and publications in general education journals could suggest a position that engineering education is part of the broader field of education practice, rather than a separate silo.

Assessment life-cycle element, or simply *assessment element*, was coded for. This is the specific assessment practice that literature focuses on (e.g. moderation or feedback). The term *element* is used as assessment covers a wide range of complex procedures, practices and activities, each of these can be thought of as a specific element relating to a specific stage of a life-cycle. Articles were coded based on the main focus of the paper, usually indicated in the title, abstract, research questions and conclusion. Naturally, many papers discussed a range of assessment elements throughout, but these were typically not the focus. For example, an article covering a workshop for developing students understanding of marking rubrics would be coded as exploring the assessment element of student understanding of assessment (AE 15). While, the article may also discuss evaluative criteria (AE 7), along with the design of university-based (AE 1) and formative assessment (AE 4), the primary focus is student understanding. The majority of literature reviewed had a main focus of one ($n = 46$) or two ($n = 39$) assessment elements (literature with a focus of three assessment elements, $n = 14$; four element, $n = 3$; five elements, $n = 1$).

From a theoretical and practical position, extracting relevant data was based on the large body of work on assessment in higher education. Each study was coded for which assessment element is focused on. Coding the sources in such a way allowed for evaluation of how current practice in engineering education compares to broader trends in higher education. Therefore, if certain elements are absent in engineering specific literature, the more general HE literature could provide a starting point for further investigation.

Finally, codes were used to review *article method*. Table 2 details the coding scheme used to summarise and evaluate the overarching research approach used in the studies. Codes are used to represent the overarching forms of research on educational practice. These included non-empirical literature proposing new theoretical positions (theoretical articles); literature reviews to establish best or current practice (review papers); literature describing of the development of materials/resources or the implementation of new approaches (accounts); literature reporting efficacy tests of materials, resources or new teaching approaches (co-relational methods, i.e. statistical comparison of two or more measures); and literature detailing robust evaluations of materials, resources or new teaching approaches with comparison of different participant conditions (experimental designs). While this list of methods does move from focusing on more qualitative data to focusing on more quantitative data, it is not proposed that the order represents any form of hierarchy. In this review all papers contribute equally, it is merely that different methods provide different forms of evidence for enhancing assessment. For instance, experimental designs can provide robust evidence for the efficacy of assessment approaches, however, are difficult to conduct in practice (see Shadish, Cook, and Campbell 2002, for different experimental designs). Furthermore, while case studies are at

Table 2. Article method coding scheme.

Method	Summary	Pro	Cons
Theoretical	Article details theoretical model or opinion/ selective literature review	Targets specific area in detail	Theoretical or unevaluated
Review	Article is a form of literature review (e.g. systematic or meta-review), it contains methods for selection of literature	Reviews vast amount of literature	Rarity
Accounts	Article details the opinion of students/staff, or a summary of/report on practice (e.g. case studies)	Contextualises methods with target population	Lack of robust evaluation of method
Co-relational	Article evaluates an assessment method or instrument by comparing two or more measures, usually statistically	Evaluates learning outcomes	Lack of control/ comparison group
Experimental	Article evaluates an assessment method or instrument using experimental methods, involving assigning participants to different conditions	Robust evaluation of learning outcomes, with control/ comparison	Often lacks contextual information about participants

times criticised for providing simple narrative descriptions of practice, they actually provide detailed accounts of how methods and theory are implemented in educational institutions. Such descriptions can be extremely beneficial for teaching staff wishing to apply new assessment methods.

2.5 Data analysis

As the main aim was to provide an overview of the current work being conducted on assessment in engineering education, the data analysis in this paper is limited to reporting descriptive statistics about the content of the articles. Section 3 summarises the articles in relation to the coding categories, note that all percentages are rounded to the nearest whole number. Where appropriate, examples from the articles are provided and referenced. Furthermore, the range of topics explored in the included articles was considerable. However, as Figure 1 showed, there were no articles that were sufficiently similar to support conducting a meta-analysis. That is, there were not enough papers looking at the same area of assessment, with statistical measures reported, to allow for a statistical meta-analysis. Finally, comparisons are conducted to explore whether journal type (engineering education journal or general higher education journal) and author type (engineering educators, non-engineers and mixed author teams) influence the assessment element studies and the article method used.

In this study, there were no recognised ethical issues, due to the focus on secondary data analysis.

3. Results

3.1 Author type

In 33% of the 104 articles all authors were from an engineering department, in 26% articles no authors were from engineering (e.g. academics working in a school of education, a central academic development team or in libraries) and in 41% articles there was a mixture of engineers and non-engineers as authors. There is a possibility that some authors may have been engineers and since taken an appointment in a different department. However, this would mean that their primary focus is no longer engineering education.

3.2 Journal type

Of the 104 articles, 77% articles were published in engineering specific journals (e.g. *Journal of Engineering Education*), 23% were from broader higher education journals (e.g. *Studies in Higher Education*).

3.3 Assessment life-cycle elements

Consideration of assessment life-cycle elements, or assessment elements, was central to the aim of this literature review, in that the analysis aimed to explore which of the various assessment processes were commonly explored in engineering education and which may have received little attention. As detailed in the methods, the articles were coded using a framework that considered 21 different assessment sub-areas or *assessment elements* (see Table 1). Table 3 (column *n*) details the number of articles that related to the different assessment elements. This table shows the major focus of the articles, typically, articles had one major assessment focus, however, some articles covered more than one assessment element.

As Table 3 shows a considerable proportion of the articles covered elements around instrument-level design (the development of individual modular assessments), including university-based assessments (e.g. exams, coursework, laboratory practices), assessment of professional skills (e.g. group work), formative assessment (e.g. peer- or self-assessment) and e-design (e.g. virtual labs). A secondary major literature area was programme level qualitative load, specifically work on curriculum mapping for accreditation purposes. Finally, literature on marking practices represented the third major literature focus, which was shown through the large number of articles that focused on the assessment elements of marking, feedback construction, post-marking moderation and inter- and intra-student validity.

Table 3. Assessment elements covered by articles, split by journal and author type.

			Percentage of articles				
			Split by journal type		Split by author type		
Code	Assessment element	*n*	Non-eng. ed. journal	Eng. ed. journal	Non-eng. author/s	Mixed author team	Eng. ed. author/s
Instrument level design							
AE 1	University-based	23	8%	13%	4%	16%	14%
AE 2	Workplace-based	2	-	1%	2%	1%	-
AE 3	Professional skills	25	8%	15%	14%	16%	9%
AE 4	Formative assessment	21	14%	11%	10%	15%	8%
AE 5	e-design	16	8%	9%	10%	8%	8%
Sub-total			39%	48%	41%	57%	39%
Programme level design							
AE 6	Constructive alignment	2	-	1%	2%	1%	-
AE 7	Evaluative criteria	2	3%	1%	2%	1%	-
AE 8	Quantitative load	4	2%	2%	4%	1%	2%
AE 9	Chronologic load	2	-	1%	-	1%	2%
AE 10	Qualitative load	20	8%	11%	12%	10%	11%
AE 11	Performance standards	1	-	1%	-	-	2%
AE 12	Administrative	4	3%	2%	-	1%	5%
AE 13	Assessment culture	4	8%	1%	4%	3%	-
Sub-total			25%	20%	25%	19%	20%
Communication (support and engagement)							
AE 14	Communication about assessment	3	6%	1%	-	3%	2%
AE 15	Student understanding of assessment	4	6%	1%	2%	3%	2%
Sub-total			11%	2%	2%	5%	3%
Marking							
AE 16	Pre-marker training	1	-	1%	2%	-	-
AE 17	Marking	8	3%	6%	2%	3%	11%
AE 18	Feedback construction	11	3%	7%	8%	3%	8%
AE 19	Post-marking moderation	9	-	6%	6%	4%	5%
Sub-total			6%	19%	18%	10%	23%
Review (of task, rubrics and results)							
AE 20	Inter and intra student validity	19	11%	10%	10%	8%	13%
AE 21	Rationale for assessment task	4	8%	1%	4%	1%	2%
Sub-total			19%	11%	14%	9%	14%

Assessment elements that received less attention in the articles reviewed included literature on workplace-based assessment (e.g. assessment of students on placements in industry settings), all other programme level design elements (including constructive alignment, chronological load or enhancing the assessment culture), and issues related to student engagement (e.g. communication about assessments to students and developing students' understanding of assessment).

Table 3 details the assessment elements explored in articles, split by journal type (general higher education journals and engineering education-specific journals), as indicated by the two columns *Non-eng ed. journal* and *Eng. ed journal*. These columns show the percentage of articles focusing on a specific assessment element in relation to all the articles reviewed from the journal type. For example, the table shows that in engineering education-specific journals 13% of articles focus on assessment element 1 (AE 1) and 25% of articles in non-engineering education journals explore assessment elements related to programme level design.

Also shown in Table 3 are the assessment elements explored in articles, split by author type (non-engineering educators, engineering educators and mixed author teams), as indicated by the three columns *Non-eng author/s, Mixed author team* and *Eng. ed author/s*. These columns show the percentage of articles focusing on a specific assessment element in relation to all the articles reviewed from the author type category. For example, the table shows that articles with a mixed author team 16% of articles focus on assessment element 1 (AE 1) and 25% of articles from non-engineering authors explore assessment elements related to programme level design.

3.4 Approaches to the study of assessment in engineering education

A further part of mapping out the literature landscape on assessment in engineering education was the consideration of the main approach used in the studies. Table 4 details what is termed article method, the overarching approach used in articles. This is based on the coding framework in Table 2, see Section 2.4. As Table 4 shows, the majority of articles were accounts (case studies on current practice or actor reflections on a new assessment method) and co-relational studies (evaluations of new or existing practice). An example of an article using an accounts-based method would be Michelsen et al. (2017); an example of co-relational method includes Uribe et al. (2016). There were a reasonable number of experimental studies (e.g. Jaeger and Adair 2015) and theoretical contributions about assessment practices (e.g. Mutereko 2018). Finally, as expected, literature reviews such as this one were absent in the dataset.

Table 4 also details the article method split by journal type (general higher education journals and engineering education-specific journals), as indicated by the two columns *Non-eng ed. journal* and *Eng. ed journal*. These columns show the percentage of articles using a specific assessment element in relation to all the articles reviewed from that journal type. A Pearson's Chi-square test for independence was conducted comparing article method (theory, accounts, correlational and experimental) with journal type (non-eng ed. and eng ed. journals). There was no significant interaction between the two variables, X^2 (3) = 3.29, $p = 0.348$.

Also covered in Table 4 is the article method split by author type (non-engineering educators, engineering educators and mixed author teams), as indicated by the three columns *Non-eng*

Table 4. Methods used in articles, split by journal and author type.

		Percentage of articles				
		Split by journal type		Split by author type		
Method	*n*	Non-eng. ed. journal	Eng. ed. journal	Non-eng. author/s	Mixed author teams	Eng. ed. author/s
Theoretical	11	1%	9%	8%	3%	6%
Review	0	-	-	-	-	-
Accounts	31	10%	20%	7%	11%	13%
Co-relational	48	11%	36%	14%	22%	10%
Experimental	14	2%	12%	3%	6%	5%

author/s, Mixed author teams and *Eng. ed author/s*. These columns show the percentage of articles using different methods in relation to all the articles reviewed from the author type category. A Pearson's Chi-square test for independence was conducted comparing article method (theory, accounts, correlational and experimental) with author type (non-eng. ed. authors, eng. ed. authors and mixed author teams). There was no significant interaction between the two variables, X^2 (3) = 6.83, $p = 0.336$.

3.5 Article references

The references lists of all articles were reviewed for the number of references to engineering specific sources (those which made specific reference to 'engineering' in the title or journal name) and the number of references to non-engineering specific sources. This measure allows for the comparison of the proportion of engineering education specific references used by articles in relation to other review variables. The results serve as a proxy for establishing the extent to which engineering specific studies are integrating ideas from research on assessment in other fields. Table 5 shows the split of engineering and non-engineering specific references across the whole sample of 104 articles.

Table 6 details the number of engineering and non-engineering specific references split by journal type (non-eng. ed. and eng. ed. journals). A Mann–Whitney test was conducted to explore whether articles published in different journal types have a higher number of engineering-specific references. The results show that there is a significant difference in journal type, $U = 473.50$, $p < .001$, with articles published in non-engineering specific journals having a lower mean proportion of engineering specific references ($M = 0.20$, $SD = 0.19$) compared to articles published in engineering specific journals ($M = 0.25$, $SD = 0.38$).

Table 7 details the number of engineering and non-engineering specific references split by author type (non-eng. ed. author/s, eng. ed. author/s and mixed author teams). A Kruskal–Wallis test was conducted to explore whether articles published by different author types have differing proportions of engineering-specific references. The results show that there is a significant difference between author types, $H = 16.39$, $p < .001$, with articles published by non-engineering authors having the lowest proportion of engineering specific references ($M = 0.19$, $SD = 0.16$), articles published by engineering educators having the highest proportion ($M = 0.45$, $SD = 0.28$) and mixed author teams in the middle ($M = 0.37$, $SD = 0.22$).

4. Discussion

This mapping of the field of assessment in engineering education has shown where literature is plentiful (instrument-level design, accreditation, and marking and feedback) and where it is lacking (workplace-based assessment, student engagement and programme level design) by contrasting the field a large sample of studies to main comprehensive frameworks on assessment (e.g. the EAT framework, Evans 2016). An excellent engineering education-specific resource is *The Assessment of Learning in Engineering Education* (Heywood 2016). Heywood explores 'the validity of teaching and judging learning' in an engineering education context, outlining how research should inform practice. This book is not reviewed here as the aim was to identify current trends in assessment in engineering education research whilst highlighting areas for further investigation. Such areas for further

Table 5. Article reference type.

	No. of references	Mean	*SD*	Median	Min.	Max.
Total references	4165	40.04	24.38	35	7	177
Eng. ed. references	1297	12.47	9.46	10	0	45
Non-eng. ed. references	2868	27.58	22.44	25	1	162

Table 6. Article reference type split by journal type.

Reference type	Journal type	No. of articles	Number of references M	SD	Median	Min	Max	Total refs
Total references	Non-eng. ed.	24	41.12	17.89	39.5	7	97	987
	Eng. ed.	80	39.73	26.11	33	8	177	3178
Eng. ed. references	Non-eng. ed.	24	7.88	6.99	6	0	23	189
	Eng. ed.	80	13.85	9.71	12	0	45	1108
Non-eng. ed. references	Non-eng. ed.	24	33.25	17.61	31	6	92	798
	Eng. ed.	80	23.54	23.54	20.5	1	162	1883

exploration are suggested in relation to current work on assessment that is not discipline-specific, as ultimately general assessment practices are not specific to one field. The following discussion explores key findings and issues raised from this literature review.

4.1 Access to articles

Whilst this literature has attempted to map the current landscape of literature on assessment in engineering education, there was one main limitation to this review, which could also be considered a limitation of the current publishing landscape. Out of the 308 articles which met the inclusion criteria, only 104 could be accessed, through a subscription at a research-intensive UK University and additional contact with authors. For recently published inaccessible articles they would have been within the embargo placed by publishers (e.g. for one year after publication articles cannot be posted online) and thus pre-print versions could be made freely available on online repositories (e.g. *ResearchGate*). While many articles could not be accessed due to paywalls or absence of freely available pre-prints, there may have been patterns of missing publications (e.g. inaccessible conference proceedings) that could not be observed in our dataset. Ultimately, considering almost two-third of all articles could not be accessed, greater public access to articles would lead to greater research impact, for both theory and practice. Therefore, the first recommendation is a move to ensure easier access to articles in engineering education, either through using repositories to share pre-print versions or a move to open-access where possible. Such a change is already underway within the European Union due to *Plan S*, an open-access publishing initiative for all funded research.

4.2 Assessment-related focus of articles

4.2.1 Assessment elements with high representation

Article focus was explored by considering the assessment life-cycle element, or simply assessment element, studied (see section 2.4). Looking at the sample as a whole, a large proportion of papers focused on the design of individual or modular assessments. These articles covered various

Table 7. Article reference type split by author type.

Reference type	Author type	No. of articles	Number of references M	SD	Median	Min	Max	Total refs
Total references	Non-eng. ed.	27	52.07	33.89	41	7	177	1406
	Mixed team	43	37.56	18.50	34	10	91	1615
	Eng. ed.	34	33.65	18.62	30	8	95	1144
Eng. ed. references	Non-eng. ed.	27	10.37	10.39	6	0	43	280
	Mixed team	43	12.60	7.91	12	2	31	542
	Eng. ed.	34	13.97	10.45	12.5	0	45	475
Non-eng. ed. references	Non-eng. ed.	27	41.7	30.58	38	7	162	1126
	Mixed team	43	24.95	16.39	23	1	69	1073
	Eng. ed.	34	19.68	16.05	17	1	70	669

instrument-level assessment elements (e.g. the design of university-based assessments, AE 1 or assessment of professional skills AE 3). Such a finding is to be expected as many articles were reports on the design and implementation of innovative assessment approaches, which were trialled in the context of a single module. Ensuring assessment is well designed and fit for purpose is a key tenet of the HEA framework (HEA 2018) and a core dimension of the EAT framework (Evans 2016).

A key aspect in most current frameworks for assessment practice is programme level design. In the sampled literature a major assessment element explored was programme level qualitative load (AE 10). This covers how assessment across a programme of study is designed to address programme level learning outcomes (i.e. constructive alignment). However, in the coding scheme used, qualitative load also incorporated literature on accreditation. It is this second element that these articles focus on: how assessments across a programme of study map to learning outcomes set by accrediting bodies (e.g. ABET). Literature focusing on this element reflects the importance of accreditation procedures and the diverse nature of regulatory bodies in engineering education. Literature on professional accreditation typically focuses on reactive methods for mapping existing assessment onto accreditation criteria (i.e. curriculum mapping), rather than the proactive creation of a programme level assessment structure that is based on accreditation criteria (which could be considered as accreditation learning objectives). This disparity suggests that, although accrediting bodies have great potential to drive improvements in assessment, they can also impede progress. It is therefore important that not only engineering educators but also accrediting bodies and accreditation panels are familiar with the latest developments in assessment.

4.2.2 Assessment elements with low representation

Workplace-based assessment (AE 2), evaluating students on industry placements, was one assessment element that has received little attention in the reviewed literature, which is surprising considering the number of engineering programmes with student placements. It could be that the lack of articles on workplace-based assessment reflects a trend for universities to outsource the assessment of students on placement to placement providers. However, when assessment is outsourced it is even more important to be clear about what is assessed and how to ensure a consistent framework and it is alarming that the evidence for this is lacking. This important assessment context may provide a fruitful area for future study, which could draw on established workplace-based assessment methods in Medical Education, Teacher Training and Veterinary Education, within a framework of social science theory, as advocated by Trevelyan (2019). Specific challenges with workplace-based assessment in the engineering education context include quality assurance (e.g. are all graders applying criteria equivalently), combining numerical and non-numerical evaluations (e.g. awarding an appropriate grade for demonstrating competencies) and satisfying perceived requirements for academic content.

Further assessment elements that received little attention in the sampled literature include the majority of elements associated with programme level design. Programme level assessment refers to decisions made and actions taken at the programme level, rather than decisions about individual assignments (e.g. constructive alignment, see Biggs and Tang 2011). It provides structure, guidance and consistency for both staff and students across a programme of study. However, it is a complicated aspect of assessment design and one that is, at times, neglected. The lack of articles covering such programme level elements (i.e. constructive alignment (AE 6), evaluative criteria (AE 7), quantitative load (AE 8), chronologic load (AE 9), performance standards (AE 11), administrative processes (AE 12) and assessment culture (AE 13)) is worrying as it is a core element of many assessment frameworks (e.g. see HEA 2018; Evans 2016). All of these elements are vital to implementing a successful assessment strategy across a programme of study, as, ultimately, they influence what happens at a modular level. This is especially so in Engineering, where accreditation requirements are usually specified at programme level.

Looking outside of engineering education, there is a range of methods for evaluating assessment at a programme level (e.g. see TESTA 2019; Tomás and Jessop 2018) and a variety of frameworks to support programme level assessment enhancement (e.g. see Evans 2016; Tomás 2017). It is acknowledged that institutions may be using such methods and toolkits but not reporting their influence in the literature. That is, articles may not report the theoretical framework that informed assessment design decisions. Furthermore, the lack of literature on programme level assessment is no overt criticism of engineering education, a focus on the range of programme level elements is not a major theme in the broader higher education literature. However, it may be that engineering education is a step behind the more discipline-general literature which places a greater emphasis on programme level assessment.

One assessment element which has received much attention in the main assessment literature is communication (AE 14) and student engagement (AE 15) in advance of assessment. This is thought to encompass communication and active engagement (Evans 2016; Winstone et al. 2017), often termed evaluative judgement (Boud et al. 2018). However, this review suggests there is a lack of studies in engineering education related to the two assessment elements concerning communication (see Table 1, AE 14 and AE 15). Firstly, communication about assessment to students (AE 14). There were a limited number of articles covering how students are informed about assessment procedures, criteria, timelines, etc. Typically these are standardised activities where students are informed on how to submit assessments via modules outlines or course handbooks. Secondly, there was a lack of articles on developing student understanding of assessment (AE 15), which includes work on student engagement, students' assessment literacy and developing evaluative judgement skills. This is a concern; student engagement activities are vital to avoid misunderstanding about assessment and ensure students have clear knowledge of expectations. Similar concerns have been raised by Goldsmith and Willey (2016) who found that supporting students in writing for assessments is often not part of the curriculum or considered part of engineering educators' remit. Again, literature outside of engineering education provides advice on student engagement activities (Evans 2016) and shows that engagement activities lead to improved understanding of module content (e.g. see Boud et al. 2018; Panadero and Broadbent 2018; Winstone et al. 2017) and assessment criteria (e.g. see Brignell et al. 2019; Panadero and Jönsson 2013; Panadero and Romero 2014).

4.2.3 Assessment focus interaction with other factors

Analysis was conducted to explore whether article focus, in terms of assessment element, differed depending upon (a) journal type and (b) author type. Splitting individual assessment elements by these variables is problematic, as drawing inferences is contentious for assessment elements for which there were a small number of articles. Therefore, the following discusses these variables in relation to the broader assessment areas covered in Table 1 (instrument level design, programme level design, communication, marking and review).

In non-engineering education journals (e.g. the International Journal of Science Education) there was a greater focus on assessment elements related to programme level design (AE 6-13), assessment communication (AE 14-15) and review of assessment practices (AE 20-21). Whereas, articles published in engineering education-specific journals show a greater focus on instrument-level design and marking practices. While statistical analysis could not be conducted to explore if these were significant differences (due to violation of the assumption of independence – articles could be coded for multiple assessment elements), the differences between the two journal types were considerable (see Table 4). It is recommended that engineering education practitioners seeking information on assessment consult non-engineering specific journals for all types of assessment practice, as there was still a considerable amount of research in non-engineering specific journals. Although exploring such journals would be especially beneficial when researching programme level design, communication and review practices. These findings also raise the question of why these differences exist, are they caused by author preference when submitting articles for publication, by journal publishing trends or by other factors?

In terms of influence of author type, the results show that mixed author teams are more likely to focus on assessment elements related to instrument level design and communication, non-engineering authors are more likely to focus on programme level design (AE 1-5) and review of assessment practices (AE 20-21), and engineering educators are more likely to focus on marking practices (AE 16-19). Again, statistical analyses could not be conducted. Considering the broader assessment element categories (e.g. marking) there is little difference in article focus for teams of non-engineers compared to teams of engineering educators. The exception being a greater focus on programme level design for non-engineering education teams. Furthermore, teams of engineers and teams of non-engineers appear to differ in relation to articles produced by mixed author teams. That is, mixed author teams show a greater tendency to focus topics related to instrument level design and communication, and are less likely to focus on marking and review practices. This could be caused by engineering educators collaborating with academic development teams to design new and rigorous assessment instruments or methods of student engagement, in order to drive the field of engineering education forwards (Olds, Moskal, and Miller 2005). Further research is needed to explore the situations which give rise to interdepartmental teams and whether such partnerships facilitate effective collaboration.

4.3 Commentary on article method

Articles reviewed used a range of approaches to study assessment in engineering education, ranging from theoretical position papers (i.e. where assessment elements are outlined and reflected on) through accounts-based methods (e.g. case studies) and co-relational studies (e.g. comparing student performance before and after an intervention) to experimental designs (i.e. those that involve different assessment 'treatments'), see Section 2.4 for further details. It is stressed that this is not a hierarchy, each method plays a vital part in developing assessment in engineering education. The following conclusions about article method hold regardless of whether the article was published in an engineering specific journal or a general higher education journal, and regardless of the authors of the article (non-engineering educators, engineering educators or mixed author teams).

This review suggests that the majority of engineering education assessment literature uses accounts-based or co-relational methods (e.g. surveys, interviews and focus groups), which mirrors findings about research methods used across the entire field of engineering education (Olds, Moskal, and Miller 2005), For a range of article topics, both methods are relatively easy to employ, it is not overly difficult to gather student reflections on a new assessment method (i.e. accounts) nor to measure student performance before and after an intervention (i.e. co-relational). This is not to claim any results are, therefore, unsophisticated or uninformative. Both approaches can be used to collect and analyse complex datasets, and both can lead to results that have profound implications. Case studies, in particular, can provide a wealth of information on how to incorporate theory from the wider literature into assessment design and what problems may be faced when implementing certain assessment methods. However, there reaches a point where additional descriptive case studies have little effect on the advancement of a field, unless, that is, they follow a period of major change and innovation.

Experimental designs differ from co-relational methods in that when evaluating an assessment method they include some form of participant allocation, between the main treatment and either an alternative treatment or control group. Such studies provide more robust evaluations of assessment methods, however, can involve considerable time costs. It is suggested that such benefits outweigh the costs. A move to more experimental style evaluations would expand the evidence base on assessment in engineering education and could result in more evidence-led practice (for support see Shadish, Cook, and Campbell 2002).

The final group of articles focused on the development and dissemination of theoretical ideas about assessment in engineering education. Such papers would suggest new assessment approaches or critically reflect on existing practices, without the inclusion of new empirical data.

Theoretical contributions undoubtedly can drive change; however, the field must ensure that such theories are tested in settings and that the results are shared, both in terms of the effectiveness of the approach and how the theory is translated into practice by teaching staff.

Systematic literature reviews are missing (e.g. a selective or systemic review of assessment in engineering education), suggesting that a systematic-based literature review such as this paper may be timely in supporting the development of the field of assessment in engineering education, especially in terms indicating areas for future systematic reviews.

4.4 Article references

Article referencing patterns were explored considering the number of engineering specific references and the number of non-engineering specific references in each paper. Across the entire sample, 31% of references were to engineering specific sources, suggesting that assessment in engineering education does draw on the wider literature base. Although there are limitations to this finding as sources were coded based on the information presented in article reference lists, the actual sources were not reviewed.

Considering article referencing patterns by journal type shows a significant difference in the proportion of engineering specific references used in different journal types. 35% of references in engineering specific journals were to engineering specific sources, whereas in non-engineering specific journals this decreased to 19%, confirming engineering specific publications tend to rely more greatly on work in the same field. This is supported by the analysis of article references in relation to author type, where a significant difference was found for referencing patterns based on author type. References in 42% of articles authored by engineering educators were engineering specific sources, this proportion decreased to 34% of references when the articles were authored by mixed teams and decreased further still to 20% of references when authored by non-engineering authors. A possible explanation for these differences is that larger interdisciplinary research teams are formed by those with more experience of teaching and researching in engineering education, therefore, can draw on a wider body of knowledge. Conversely, those at an earlier career stage may be less familiar with the wider body of research and more likely to undertake smaller-scale research projects with other engineering education colleagues. Consequently, these differences may be representative of career progression. Regardless, the findings suggest that when engineering educators work with those outside engineering, be that publishing in general HE journals or working in interdisciplinary teams, the resulting work is more likely to be influenced by a wider range of theories, concepts and practice. It is not proposed that work reported in engineering journals by engineering educators is in any way less important. Rather this analysis suggests that practitioners and researchers should consult with those outside of engineering, to avoid the potential of cyclical referencing and engineering education becoming an echo chamber for work on assessment. This could be achieved through reading and publishing in general HE journals or through conducting research and assessment design with non-engineers (e.g. those working schools of education, in academic development teams or in libraries).

4.5 Limitations

The main limitation was the inability to access 204 of the 308 papers that met the inclusion criteria. As discussed in section 4.1, access issues were primarily due to pay-walled articles along with difficulty finding freely available pre-prints or articles shared after journal embargos. An additional limitation may relate to the search strategy used along with databases searched, conference proceedings are often underrepresented in databases compared with journal articles. Therefore, the strategy used may have missed such publications on assessment in engineering education. These limitations mean that it was not possible to capture the complete picture of current research. There may be patterns of publication that could considerably alter the findings presented here.

For example, a large proportion of the inaccessible literature may focus on programme level design, rendering the suggestion that a greater focus in this area is needed a moot point. While acknowledging that there may be patterns of publication missed in this literature review that may have resulted in different findings, recommendations have been made based on the 104 articles that could be accessed and represent the literature that practitioners at a research-intensive university could access.

Our assessment framework and elements provided a lens from which to explore the engineering literature. It is important to acknowledge, that our review is limited to the scope of the framework. Nevertheless, the framework is forward-looking in that it captures key recommendations from the literature for best practice and framed our inquiry to explore these.

4.6 Conclusions

Reported here are the results of a systematic-based literature review into assessment in engineering education. The focus was on trends in the current literature base (including a focus on the design of individual assessments, accreditation and marking practices) and highlighted areas for further investigation which could benefit from additional work (including workplace-based assessment, student engagement and programme level design). This review is intended to inspire others to conduct systematic reviews in other areas of engineering education to further advance the field.

In terms of recommendations, it is argued that a greater focus is required in several areas:

- Workplace-based assessment;
- Student engagement;
- Programme level design;
- Systematic comparisons of assessment approaches through experimental designs, to complement and extend existing work using theoretical, accounts and co-relation methods.
- Interdisciplinary work conducted with non-engineering educators (e.g. academic development teams or librarians) to gain different insights into assessment practices and engage with the wider field of education.

These recommendations are specific for engineering education, as in the broader higher education literature there exists a vast amount of interdisciplinary work that seeks to identify and evaluate best assessment practices (see section 1 for summaries of research on various assessment elements).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Dr Jonathan G. Halls is a British Academy Postdoctoral Fellow at the Learning Sciences Research Institute, School of Education, University of Nottingham. His work focuses on preconceptions that individuals bring to learning situations, specifically those that can be problematic in the context of science education. He investigates a) how these preconceptions are culturally influenced and b) how they can be challenged by, and incorporated into the design of, educational materials and learning environments. Dr Halls has also worked on a number of projects related to assessment in higher education.

Dr Carmen Tomás currently works as Associate Director of Educational Excellence in the University of Nottingham. Carmen leads institutional research in practices such as curriculum design, student engagement in assessment and review methods on assessment. She leads a programme of work addressing University practices on curriculum design, assessment and technologies. Topics of interest include: programme level designs and load; rubric design;

student engagement in assessment for the development of evaluative judgement; marking and moderation practices. Currently exploring the use of machine learning analyses with assessment data.

Dr John S. Owen joined the University of Nottingham as a lecturer in the Department of Civil Engineering in 1993, having completed his BEng (Hons) and PhD in Civil Engineering at the University of Bristol. He was subsequently promoted to Senior Lecturer, now Associate Professor, and served as Head of the Department of Civil Engineering from 2009-2014. Dr Owen is a past chairman of the UK Wind Engineering Society (2009-2012) and was chair of the 6th European and African Conference on Wind Engineering held in Cambridge in 2013. He was chair of the Association of Civil Engineering Departments for 2014-15 and was the ACED representative on the Engineering Professors' Council. Dr Owen is a Fellow of the Institution of Civil Engineers, a Senior Fellow of the Higher Education Academy, a Member of the Institute of Physics and a Chartered Engineer. In 2016 he was elected a Fellow of the Wind Engineering Society. Dr Owen is also a member of the Centre for Structural Engineering & Informatics Research Group.

Professor Kamel Hawwash is Professor in the School of Civil Engineering and Immediate Past President of the European Society for Engineering Education, SEFI. He is the Convenor of the long-established MSc Construction Management Programme. He was Head of the School of Civil Engineering from 2008 to 2011. Between 2009 and 2012 he was the Regional Director for the Midlands and East Anglia for the National HE STEM Programme. Professor Hawwash is past Chair of the Association of Civil Engineering Departments, and member of the Engineering Education, Employability and Skills sub-committee of the Engineering Professors' Council and Associate of the Engineering Council, UK.

ORCID

Jonathan G. Halls http://orcid.org/0000-0002-0945-5891
Carmen Tomás http://orcid.org/0000-0001-6163-2907
John S. Owen http://orcid.org/0000-0002-5731-1577
Kamel Hawwash http://orcid.org/0000-0002-5642-4334

References

Baartman, L. K. J., T. J. Bastiaens, P. A. Kirschner, and C. P. M. Van der Vleuten. 2007. "Evaluation Assessment Quality in Competence-Based Education: A Qualitative Comparison of Two Frameworks." *Educational Research Review* 2: 114–129.

Bearman, M., P. Dawson, D. Boud, S. Bennett, M. Hall, and E. Molloy. 2016. "Support for Assessment Practice: Developing the Assessment Design Decisions Framework." *Teaching in Higher Education*. doi: 10.1080/13562517.2016.1160217.

Biggs, John, and C. Tang. 2011. *Teaching for Quality Learning at University*. 4th ed. Buckingham: Open University Press.

Bloxham, Susan. 2014. "Assessing Assessment: New Developments in Assessment Design, Feedback Practices and Marking in Higher Education." In *A Handbook for Teaching and Learning in Higher Education: Enhancing Academic Practice*, edited by Heather Fry, Steve Ketteridge, and Stephanie Marshall, 4th ed., 107–122. Abingdon: Taylor & Francis (Routledge).

Bloxham, S., and P. Boyd. 2007. *Developing Effective Assessment in Higher Education: A Practical Guide*. London: McGraw-Hill Education.

Bloxham, Susan, Claire Hughes, and Lenore Adie. 2016. "What's The Point of Moderation? A Discussion of The Purposes Achieved Through Contemporary Moderation Practices." *Assessment and Evaluation In Higher Education* 41 (4): 638–653.

Boud, D., R. Ajjawi, P. Dawson, and J. Tai. 2018. *Developing Evaluative Judgment in Higher Education*. Abingdon Oxon: Routledge.

Brame, C. 2013. "Writing Good Multiple Choice Test Questions." https://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions/.

Brignell, C., T. Wicks, C. Tomás, and J. Halls. 2019. "The Impact of Peer Assessment on Mathematics Students' Understanding of Marking Criteria and Their Ability to Self-Regulate Learning." *MSOR Connections* 18 (1): 46–55.

Brookhart, S. M. 2018. "'Appropriate Criteria: Key to Effective Rubrics'." *Frontiers in Education: Assessment, Testing and Applied Measurement* 3 (22): 1–12.

Brown, George. 2001. *Assessment: A Guide For Lecturers*. Liverpool: LTSN Generic Centre.

Cockett, A., and C. Jackson. 2018. "The use of Assessment Rubrics to Enhance Feedback in Higher Education: An Integrative Literature Review." *Nurse Education Today* 69: 8–13.

Davies, P. 2000. "The Relevance of Systematic Reviews to Educational Policy and Practice." *Oxford Review of Education* 26 (3–4): 365–378.

Dochy, Filip. 2009. "The Edumetric Quality of New Modes of Assessment: Some Issues and Prospects." In *Assessment, Learning and Judgement in Higher Education*, edited by Gordon Joughin, 1–30. Dordrecht: Springer.

Dochy, F., M. Segers, and D. Sluijsmans. 1999. "The Use of Self-, Peer- and Co-assessment in Higher Education: A Review." *Studies in Higher Education* 24 (3): 331–350.

Evans, C. 2016. *Enhancing Assessment Feedback Practice in Higher Education: The EAT Framework*. Southampton: University of Southampton.
Finelli, C. J., M. Borrego, and G. Rasoulifar. 2016. "Development of a Taxonomy of Keywords for Engineering Education Research." *IEEE Transactions on Education* 56 (4): 219–241.
Galvin, A., E. Noonan, and G. O'Neill. 2012. *Assessment Workload and Equivalences*. Dublin: University College of Dublin. https://www.ucd.ie/t4cms/UCDTLA0038.pdf.
Glass, G. V. 2016. "One Hundred Years of Research: Prudent Approaches." *Educational Researcher* 46 (2): 69–72.
Goldsmith, R., and K. Willey. 2016. "'It's not my job to Teach Writing': Activity Theory Analysis of [Invisible] Writing Practices in the Engineering Curriculum." *Journal of Academic Language & Learning* 10 (1): 118–129.
Gough, D. 2007. "Weight of Evidence: A Framework for the Appraisal of the Quality and Relevance of Evidence." *Applied and Practice-Based Research* 22 (2): 213–228.
Hartley, P., and R. Whitfield. 2012. *Programme Assessment Strategies (PASS)*. York: The Higher Education Academy.
HEA. 2012. "A Marked Improvement." https://www.heacademy.ac.uk/knowledge-hub/marked-improvement.
HEA. 2018. "Framework for Transforming Assessment in Higher Education." https://www.heacademy.ac.uk/knowledge-hub/framework-transforming-assessment-higher-education.
Heywood, J. 2016. *The Assessment of Learning in Engineering Education: Practice and Policy*. London: Wiley-IEEE Press.
Jackel, B., J. Pearce, and A. Radloff. 2017. *Assessment and Feedback in Higher Education: A Literature Review for the Higher Education Academy*. York: the Higher Education Academy. https://s3.eu-west-2.amazonaws.com/assets.creode.advancehe-document-manager/documents/hea/private/hub/download/acer_assessment_1568037358.pdf.
Jaeger, M., and D. Adair. 2015. "Using an Evidential Reasoning Approach for Portfolio Assessments in a Project-Based Learning Engineering Design Course." *European Journal of Engineering Education* 40 (6): 1–15. doi: 10.1080/03043797.2014.1001817.
Jessop, T., and C. Tomás. 2017. "The Implications of Programme Assessment Patterns for Student Learning." *Assessment and Evaluation in Higher Education* 42 (6): 990–999. doi: 10.1080/02602938.2016.1217501.
Lau, A. M. S. 2015. "'Formative Good, Summative bad?' – a Review of the Dichotomy in Assessment Literature." *Journal of Further and Higher Education* 40 (4): 509–525.
Messick, Samuel. 1994. "The Interplay of Evidence and Consequences in the Validation of Performance Assessments." *Educational Researcher* 23 (2): 13–23.
Messick, Samuel. 1996. "Validity in Performance Assessments." In *Technical Issues in Large-Scale Performance Assessment*, edited by Gary W. Phillips, 1–18. Washington, DC: National Centre for Educational Statistics.
Michelsen, S., A. Vabø, H. Kvilhaugsvik, and E. Kvam. 2017. "Higher Education Learning Outcomes and Their Ambiguous Relationship to Disciplines and Professions." *European Journal of Education* 52: 56–67.
MMU. 2019. "The Assessment Lifecycle, University Teaching Academy." https://www.celt.mmu.ac.uk/assessment/index.php/.
Mutereko, S. 2018. "Analysing the Accreditation of Engineering Education in South Africa Through Foucault's Panopticon and Governmentality Lenses." *Assessment & Evaluation in Higher Education* 43 (2): 235–247. doi:10.1080/02602938.2017.1330395.
Nawaz, S., and J. Strobel. 2016. "Authorship and Content Analysis of Engineering Education Research: A Case Study." *International Journal of Engineering Pedagogy* 6 (2): 39–51.
Nicol, D. 2010. "From Monologue to Dialogue: Improving Written Feedback Processes in Mass Higher Education." *Assessment and Evaluation in Higher Education* 35 (5): 501–517.
Olds, B. M., B. M. Moskal, and R. L. Miller. 2005. "Assessment in Engineering Education: Evolution, Approaches and Future Collaborations." *Journal of Engineering Education* 94 (1): 13–25. doi:10.1002/j.2168-9830.2005.tb00826.x.
Panadero, E., and J. Broadbent.2018. "Developing Evaluative Judgement: Assessment for Knowing and Producing Quality Work." In *Developing Evaluative Judgement: Assessment for Knowing and Producing Quality Work*, edited by D. Boud, R. Ajjawi, P. Dawson, and J. Tai. Abingdon: Routledge.
Panadero, E., and A. Jönsson. 2013. "The use of Scoring Rubrics for Formative Assessment Purposes Revisited: A Review." *Educational Research Review* 9: 129–144. doi: 10.1016/j.edurev.2013.01.002.
Panadero, E., and M. Romero. 2014. "To Rubric or not to Rubric? The Effects of Self-Assessment on Self-Regulation, Performance and Self-Efficacy." *Assessment in Education: Principles, Policy and Practice* 21 (2): 133–148. doi 10.1080/0969594X.2013.877872.
Price, M., C. Rust, B. O'Donovan, H. Handley, and R. Bryant. 2012. *Assessment Literacy: The Foundation for Improving Student Learning*. Oxford: Oxford Centre for Staff and Learning Development.
Reddy, Y. M., and H. Andrade. 2010. "A Review of Rubric use in Higher Education." *Assessment & Evaluation in Higher Education* 35 (4): 435–448.
Shadish, W. R., T. D. Cook, and D. T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont: Wadsworth Publishing.
Suskie, L. 2018. *Assessing Student Learning: A Common Sense Guide*. 3rd ed. San Francisco: Jossey-Bass.
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: 467–497.

Tavakol, M., and R. Dennick. 2017. "'The Foundations of Measurement and Assessment in Medical Education." *Medical Teacher* 39 (10): 1010–1015. doi: 10.1080/0142159X.2017.1359521.

TESTA. 2019. "Transforming the Experience of Students through Assessment." https://www.testa.ac.uk/.

Todorovski, B., E. Nordal, and T. Isoski. 2015. *Overview on Student-Centered Learning in Higher Education in Europe: Research Study*. Brussels: The European Students' Union.

Tomás, C. 2017. "An Assessment Framework for Enhanced Institutional Assessment Literacy in Practice." Conference: Higher Education Academy, Engaging Students Through Assessment, At York.

Tomás, C., and T. Jessop. 2018. "Struggling and Juggling: a Comparative Study of Student Assessment Loads Across Institution Types." *Assessment and Evaluation in Higher Education*. doi: 10.1080/02602938.2018.1463355.

Tomás, C., and R. Scudamore. 2014, August. "Using an Assessment Conceptual Framework to Facilitate Institutional Transformation of Assessment." Earli Assessment and Evaluation Sig, Madrid.

Tremblay, K., D. Lalancette, and D. Roseveare. 2012. "Assessment of Higher Education Learning Outcomes – AHELO: Feasibility Study Report." OECD. oecd.org/edu/ahelo.

Trevelyan, T. 2019. "Transitioning to Engineering Practice." *European Journal of Engineering Education* 44 (6): 821–837. doi: 10.1080/03043797.2019.1681631.

Uribe, M. D. R., A. J. Magana, J. H. Bahk, and A. Shakouri. 2016. "Computer Simulations as Virtual Laboratories for Online Engineering Education: A Case Study in the Field of Thermoelectricity." *Computer Applications in Engineering Education* 24 (3): 428–442.

van der Vleuten, C. M. P., L. Schuwirth, E. W. Driessen, M. Govaerts, and S. Heeneman. 2015. "12 Tips for Programmatic Assessment." *Medical Teacher* 37: 641–646.

Winstone, N. E., R. A. Nash, M. Parker, and J. Rowntree. 2017. "Supporting Learners' Agentic Engagement with Feedback: a Systematic Review of the Literature and a Taxonomy of Recipience Processes." *Educational Psychologist* 51 (1): 17–37.