



INTERNATIONAL LABOUR ORGANIZATION  
UNITED NATIONS EDUCATIONAL, SCIENTIFIC  
AND CULTURAL ORGANIZATION



---

**CEART/13/2018/4**

---

## **Joint ILO–UNESCO Committee of Experts on the Application of the Recommendations concerning Teaching Personnel (CEART)**

### **Assessing teaching in higher education: Current issues, trends and recommendations**

Amanda Jarrell  
McGill University  
June 2018

Report submitted to the 13th Session of the Joint ILO–UNESCO Committee of Experts  
on the Application of the Recommendations concerning Teaching Personnel (CEART)

Geneva, 2018





This is an open access work distributed under the Creative Commons Attribution-ShareAlike 3.0 IGO License (<http://creativecommons.org/licenses/by-sa/3.0/igo>). Users can re-use, share, adapt and build upon the original work, even for commercial purposes, as detailed in the License. Any new works that use the original content must carry the same CC-BY-SA licence. The ILO and UNESCO must be clearly credited as the owners of the original work. The use of the emblems of the ILO and/or UNESCO is not permitted in connection with users' work.

**Translations** – In case of a translation of this work, the following disclaimer must be added along with the attribution: *This translation was not created by the ILO or UNESCO and should not be considered an official ILO or UNESCO translation. Neither ILO nor UNESCO is responsible for the content or accuracy of this translation.*

**Adaptations** – In case of an adaptation of this work, the following disclaimer must be added along with the attribution: *This is an adaptation of an original work by the ILO and UNESCO. Responsibility for the views and opinions expressed in the adaptation rests solely with the author or authors of the adaptation and are not endorsed by the ILO or UNESCO.*

All queries on rights and licensing should be addressed to ILO Publishing (Rights and Licensing), 1211 Geneva 22, Switzerland, or by email to [rights@ilo.org](mailto:rights@ilo.org).

---

*Assessing teaching in higher education: Current issues, trends and recommendations*, by  
Amanda Jarrell

This report served as a background study for the 13th session of the Joint ILO/UNESCO Committee of Experts on the Application of the Recommendations concerning Teaching Personnel (CEART), held in Geneva, 1-5 October 2018.

Also available in French: *L'évaluation de l'enseignement dans l'enseignement supérieur: problèmes, tendances et recommandations*

---

The designations employed in ILO and UNESCO publications, which are in conformity with United Nations practice, and the presentation of material therein do not imply the expression of any opinion whatsoever on the part of the ILO or UNESCO concerning the legal status of any country, area or territory or of its authorities, or concerning the delimitation of its frontiers.

The responsibility for opinions expressed in signed articles, studies and other contributions rests solely with their authors, and publication does not constitute an endorsement by the ILO or UNESCO of the opinions expressed in them.

Reference to names of firms and commercial products and processes does not imply their endorsement by the ILO or UNESCO, and any failure to mention a particular firm, commercial product or process is not a sign of disapproval.

<p>This publication was produced by the Document and Publications Production, Printing and Distribution Branch (PRODOC) of the ILO. Graphic and typographic design, layout and composition, manuscript preparation, copy editing, proofreading, printing, electronic publishing and distribution. PRODOC endeavours to use paper sourced from forests managed in an environmentally sustainable and socially responsible manner. Code: DTP-CAD-ICA</p>
--

---

## **Contents**

	<i>Page</i>
Abstract .....	1
Current trends in higher education .....	1
Introduction to assessing teaching in higher education .....	2
Defining effective teaching in higher education.....	3
Assessment and evaluation principles .....	4
Validity, reliability and fairness .....	5
Assessing teaching: Methods and metrics .....	6
Student evaluations of teaching.....	6
Peer observations of teaching.....	8
Teaching portfolios .....	9
Student learning outcomes .....	10
Emerging trends and issues .....	12
Casualization of teaching staff.....	12
E-learning and teaching online.....	13
Summary of recommendations.....	14
Areas for future research .....	15
Conclusions .....	15
References .....	17

---

## Abstract

Assessing teaching to identify opportunities for professional development and for evaluative purposes are essential for ensuring quality learning and teaching in higher education. To this end assessment systems must allow for the collection of reliable data that can be used to make valid and fair interpretations about effective teaching. Assessments of teaching must also meet the new demands and challenges created by the massification and increased application of quality assurance systems in higher education. The purpose of this paper is to discuss these challenges and issues and to propose recommendations for supporting quality assessments of teaching. In particular we discuss the challenges associated with the lack of universal standards for conceptualizing and measuring effective teaching and the affordances and limitations of current assessment methods. We also discuss emerging issues relevant for assessments of teaching including the casualization of teaching and the emergence of online teaching. Several policy recommendations are provided including the development of consistent assessment standards, access to training in measurement and evaluation and increased transparency in assessment practices.

## Current trends in higher education

In recent decades, there has been a global expansion of higher education. In 1970, the UNESCO Institute for Statistics (UIS) estimated that there were approximately 32.5 million students enrolled in higher education. In 2010, the number of students enrolled grew to roughly 178 million students and it is expected that this number will increase to nearly 263 million by 2025 (Altbach et al., 2009). This growth occurred in stages across different regions with Canada and the United States achieving mass education first in the 1960s and Western Europe and Japan in the 1980s (Altbach et al., 2009). In the last four decades, the expansion of higher education has been more noticeable in emerging regions including sub-Saharan Africa, the Arab States and Latin America and the Caribbean, with average growth rates between 6 and 8.4 per cent (Altbach et al., 2009). A more recent and phenomenal growth has been seen in the Chinese higher education system. Between 2011 and 2016, the gross enrolment ratio at the tertiary level increased from 25.29 per cent to 48.44 per cent (UIS).

This growth has resulted in a global increase in the number of institutions, public and private expenditures, greater access for students and completion rates (OECD, 2017). For example, in 1949 the Indian higher education system consisted of 27 universities and in 2006 this number grew to 361 (Altbach et al., 2009). Similarly, the number of colleges and universities in China more than doubled between 2000 and 2010, increasing from 1,041 to 2,358 institutions, where the majority of this growth occurred in the private higher education sector (Stanfield and Shimmi, 2014). Governments have also implemented policies to increase access to students. To illustrate, Ghana, Kenya, United Republic of Tanzania and Uganda have reduced admission cut-offs for women to increase access and enrolment of female students (Altbach et al., 2009). Other countries, such as the United Kingdom and the United States, might continue to seek expansion of the higher education system by broadening access or increasing the level of education within the adult population (Vincent-Lancrin, 2008). In some countries, the massification of higher education has also been accompanied by an increase in public expenditure. For example, in 2010 the Australian Government spent 1.2 per cent of the gross domestic product (GDP) on tertiary education and in 2014 this amount rose to 1.4 per cent of GDP (The World Bank Education Statistics). This is because in many countries, institutions of higher education in the public sector receive grants on the number of students enrolled or the number of students that have graduated from the institution (Santiago et al., 2008).

---

This massification in higher education has introduced questions about the quality of education provided (OECD, 2017) and prompted the development and expansion of quality assurance and accreditation agencies, elevating “quality assurance ... to the top of the policy agenda in many nations” (Altbach et al., 2009, p. viii). Implicit in the broad definition of quality assurance is the “process of establishing stakeholder confidence that provision (input, process and outcomes) fulfils expectations and measures up to threshold minimum requirements” (Harvey, 2018). For example, in the United States accreditation and standards in higher education are used to ensure education quality (National Advisory Committee on Institutional Quality and Integrity, 2011). Increasingly, institutions of higher education must provide evidence that they are reliable providers of education, to assure employers, policy-makers and other stakeholders that all students are receiving a quality education. This includes the responsibility to support different kinds of learners with diverse needs, abilities and backgrounds. Mechanisms for controlling quality vary from one jurisdiction to another. The website of the International Network for Quality Assurance Agencies in Higher Education (INQAAHE) shows 187 full members, representing 93 countries. In some countries, the control is done centrally. For example, in Australia, the federal Government’s independent agency, Tertiary Education Quality and Standards Agency (TEQSA), oversees the quality of teaching and learning in higher education. Likewise, in China, the Ministry of Education launched the Quality and Reform of Undergraduate Teaching and Learning Project to improve the quality of undergraduate teaching and learning. In the United States, regional and national agencies monitor quality.

Quality assurance frameworks in higher education require that institutions develop and demonstrate effective teaching. This means that institutions must be able to identify and measure the strengths of teaching practices and areas for improvement by evaluating the input, processes and outcomes of teaching. This presents new challenges for institutions because traditionally the effectiveness of academic staff was primarily determined by research productivity and “the higher education sector is generally not well prepared for increased accountability for educational outcomes” (Lodge and Bonsanquet, 2014, p. 5). However, both massification and accountability have rendered the assessment of teaching important and introduced the imperative that quality and transparent assessment systems be put in place. In the remaining sections of this paper we discuss issues and provide recommendations for achieving this goal.

## **Introduction to assessing teaching in higher education**

Measuring and evaluating teacher performance is crucial for ensuring and demonstrating quality education. “It is essential to know the strengths of teachers and those aspects of their practice which could be further developed. From this perspective, the institution of teacher evaluation is a vital step in the drive to improve the effectiveness of teaching and learning and raise educational standards” (OECD, 2009). Assessments of teaching can target different levels within higher education including at the level of the individual, programme and institution. Assessment at each level is vital for monitoring and maintaining overall quality education. General metrics used to assess quality teaching at the institutional level are typically input- and process-based, such as student to teacher ratios and completion time and are not strongly correlated with student learning (Salmi and Saroyan, 2007; Strang et al., 2016), while quality teaching measured at the level of the teacher is often associated with learning. Assessments at the individual level also carry the additional weight of being used for personnel decisions, such as hiring and tenure, as well as identifying opportunities for professional development. Looking at appraisals of individual teaching, or even an aggregation of individual teaching results, can provide a more accurate picture of the quality of teaching and learning. The principles discussed in this paper

---

are applicable to assessments of teaching at all levels. However, given the importance of assessments of teaching at the individual level, this will be the focus of the current paper.

Assessment of teaching in higher education is generally used for improvement and accountability functions, as well as a reference for students to make course selections. Assessment of teaching, including course ratings by students, also referred to as course evaluations, were initially introduced for improvement purposes. This type of assessment is generally formative in nature and is used to provide useful feedback to teachers on how to elevate the quality of their teaching practice as well as to identify professional development needs. Assessment of teaching for accountability is typically summative and is used as a means to render teachers responsible for the quality of their teaching. Results of this type of assessment are used for making personnel decisions such as retention, renewal, promotion and even merit-based salary increases. While similar assessment methods can be used for improvement and accountability purposes, using the same method to achieve both purposes can create tensions and raise difficult challenges.

In the following sections, we discuss issues related to assessing teaching in higher education. We begin with conceptualizations of effective teaching and how these perspectives inform the development of assessment instruments. We then offer an overview of assessment and evaluation principles and how these principles can be used to guide assessment practices. Next, we examine the strengths and limitations of current methods and metrics used to assess teaching. In the final section, we provide a brief look into some emerging trends and issues relevant for assessing teaching in higher education. We conclude with a summary of our recommendations and suggestions for future research.

## **Defining effective teaching in higher education**

Teaching in higher education is a highly complex process that broadly includes “the design of curricula, choice of content and methods, various forms of teacher–student interaction and the assessment of students” (Ramsden, 2003, p. 85). How teachers approach teaching is influenced by their conceptions of teaching and learning and their perceptions of the teaching environment (Prosser and Trigwell, 1997; Trigwell and Prosser, 1996). An important question that assessments of teaching in higher education attempt to answer is if (or to what extent) teachers are teaching effectively. To answer this question, reliable reference standards or clearly defined teaching competencies for the context of higher education are needed. This leads to more general questions about what good or effective teaching is in higher education.

There are no universally accepted standards of effective teaching and this remains a contested issue in the literature (Skelton, 2005; Kember et al., 2006). Effective teaching can be largely defined according to the teaching activities which occur before teaching (preparatory) and during teaching (delivery) that produce relevant changes in students (Abrami et al., 2007). The specific expectations and minimum requirements of effective teaching are not clearly defined. It is generally accepted that the skills and practices used by effective teachers are specific to the context in which teaching and learning occurs. Thus, there is not a single process of effective teaching; there are many. In addition, there is not a single outcome of effective teaching and the value placed on these outcomes are situation-specific. Adjustments are required to meet local needs as described by various stakeholders “including students, employers, teaching and non-teaching staff, government and funding agencies, creditors, auditors, assessors and the community at large” (Hénard, 2010, p. 87). Effective teaching then is context specific, multidimensional and consists of numerous and possibly distinct acts (Abrami et al., 2007).

Data on how institutions across different regions currently define effective teaching is difficult to obtain and a comprehensive comparative analysis of institutional standards and

---

requirements for effective teaching is presently lacking. The scholarly literature on effective teaching in higher education provides some insights into perspectives on what constitutes “good” teaching in higher education. One early guiding framework for “good” teaching in higher education comes from Chickering and Gamson’s (1987) seven principles which has been cited over 7,000 times. According to this perspective, effective teaching in undergraduate education should adhere to the following principles: encourages contact between faculty and students; develops reciprocity and cooperation among students; uses active learning techniques; gives prompt feedback; emphasizes time on task; communicates high expectations; and respects diverse talents and ways of learning. While this publication is more than 30 years old, many of these guiding principles are still advocated for today (for example, Tobin et al., 2015). Several of the principles focus on the actions or behaviours of the teacher that can foster student learning. A more recent model of teaching, proposed by Saroyan and Amundsen (2001), in addition to teaching actions, takes beliefs and knowledge into account, considering all three important elements in delivering effective teaching. This model of teaching was designed primarily for faculty improvement purposes and, as such, it highlights the importance of an alignment between its main components. Beliefs about teaching and learning can include individual beliefs about the roles and responsibilities of teachers and students within a particular instructional context and of what constitutes effective teaching and learning. Knowledge pertains to knowledge about the subject matter, pedagogy and student needs. Finally, actions in this model include the processes and outcomes of teaching. This model proposes that these elements interact with one another within a particular instructional context (for example, level, course type, discipline, etc.) and therefore the definition of “teaching effectiveness” is dynamic rather than static.

Operationalizations of effective teaching are used to inform which data will be collected and accepted to make judgements about teaching, since “once the notions of quality and of teaching have been defined, the institution is in a better position to determine appropriate instruments for appraising quality” (Hénard, 2010, p. 87). The specific indicators of effective teaching in higher education and ways to measure it in this context are still unclear. The absence of accepted standards of effective teaching has created obstacles for developing consistent assessment practices of teachers in higher education. As recommended by Strang et al. (2016) “a shared understanding must be established on the notion of ‘quality teaching’ in higher education and the goals and priorities of [higher education institutions] themselves, and how these elements must be harmonized” (p. 7). A parallel to this can be observed in the K-12 sector where many regions have established professional standards or competencies for teaching which are used to support professional training, self-evaluation, reflection and assessment (for example, *The Standards of Practice for the Teaching Profession* in the Canadian Province of Ontario). In the higher education sector, a consensus between stakeholders and policy-makers of what constitutes quality teaching and what types of evidence should be collected and accepted as indicators of effective teaching would be beneficial. With increased internationalization, there is a real need to establish that graduate outcomes are comparable and the quality of teaching provided is critical for supporting what students learn. And yet open questions remain regarding how much flexibility is required to account for the goals and needs of local contexts. For example, should institutions prioritize defining effective teaching for their particular context in terms of inputs, processes and outcomes? Should definitions aim to reflect the multidimensionality of teaching and should institutions consider making these available, particularly to relevant stakeholders in an effort to provide transparency?

## Assessment and evaluation principles

No matter whether inputs, processes or outputs are targeted, assessment of teaching must yield trustworthy results. This is fundamental because if the quality of measures used to assess teaching are poor, then the value of conclusions drawn from these assessments is limited. This has important implications for high-stake decisions including making

---

personnel decisions and conclusions about the quality of education. Guiding principles for developing and judging assessments provide insights into this key issue. One set of principles has been defined by the American Psychological Association, *Standards for Educational and Psychological Testing* (American Psychological Association, 2014). These *Standards*, in part, provide guidelines for the use of assessments intended to make decisions about employee selection, placement and promotion. In this paper, we refer to these *Standards* to inform our appraisal of the current methods used to assess teaching in higher education and to make recommendations. In the next sections, we provide a brief review of the fundamental principles discussed in the *Standards*: validity, reliability and fairness. Overall these principles allow for objective appraisals of assessment and to support accurate and trustworthy assessment results.

## Validity, reliability and fairness

The three fundamental principles of assessment and evaluation are validity, reliability and fairness. The principle of validity refers to the degree to which evidence and theory support the interpretation of assessment results for proposed uses of the assessment (American Psychological Association, 2014). In other words, validity is concerned with the soundness, trustworthiness or legitimacy of the inferences made on the basis of assessment results. In the context of assessing teaching in higher education, this equates to claims about the effectiveness or quality of teaching. The principle of reliability refers to the relative precision of assessment results. This principle becomes increasingly important as the significance of interpretations escalate. For example, if the assessment measure is being used to make decisions that are not easily reversed (for example personnel decisions, including hiring and tenure), then a high degree of reliability is desirable. The more reliable the assessment measure is the less measurement error there is. Therefore, reliability influences the amount of confidence we can have in the assessment results and, by extension, the trustworthiness of interpretations based on these results. There are many factors that can influence reliability including changes to the assessment procedure, changes to the assessment, variability between raters and differences between populations. Therefore, it is important to ascertain that an assessment of teaching designed for a particular context will remain reliable in a different context. Finally, the principle of fairness refers to an assessment that “does not advantage or disadvantage some individuals because of characteristics irrelevant to the intended construct” (American Psychological Association, 2014, p. 50), including race, ethnicity, gender, age, socio-economic status, linguistic or cultural background. If some aspects of the assessment or the assessment procedure “systematically lowers or raises scores for identifiable groups of test takers [it can result in] inappropriate score interpretations for intended uses” (American Psychological Association, 2014, p. 54). Fairness must be considered throughout all stages of developing and administering assessments and interpreting assessment results.

Reliability, validity and fairness evidence must be systematically collected and documented to assert that assessment results are accurate and that interpretations can be trusted. To make claims about the validity, there are several types of evidence that can be collected including evidence based on the content of the assessment and relations to other variables. Validity evidence based on content must demonstrate that there is alignment between the content of the assessment and the construct being measured. Evidence based on relations to other variables must demonstrate that the results from the assessment are related to other variables in the expected direction and degree. It is possible to expect that the assessment results will be closely related to other variables or unrelated to other variables. Likewise, to make claims about the reliability of the assessment evidence is needed, such as evidence that the assessment instrument can provide stable results over time. Evidence of fairness is more difficult to establish because “group differences in outcomes do not in themselves indicate that a testing application is biased or unfair” (American Psychological Association, 2014, p. 54). Any group difference in outcomes requires greater scrutiny to

---

understand the cause of this difference. As an illustrative example, the results from an assessment of teaching might indicate that senior faculty perform worse than junior faculty. It is possible that this assessment unintentionally disadvantages senior faculty and advantages junior faculty. It is also possible that these differences reflect an actual difference in the effectiveness or quality of teaching. Therefore, differences based on identifiable groups should be carefully examined and when an actual bias is identified, procedures should be put in place to fairly address this bias. The development, administration and interpretation of teaching assessments should be consistent with measurement and evaluation standards such as those outlined by the American Psychological Association or equivalent. These guiding standards provide a systematic approach to the assessments of teaching that can support quality assessment practices. Evidence of reliability, validity and fairness should be collected and documented to ensure that assessment results allow for accurate and trustworthy interpretations about the quality of teaching. Institutions should also provide documentation to guide the appropriate use and interpretation of the assessment. This documentation should include information regarding the nature of the assessment, the use for which it was developed, the process involved in developing the assessment, information relating to scoring and interpretation, and evidence of validity, fairness and reliability (American Psychological Association, 2014). This is important for both assessments developed locally and circulated within institutions and those developed externally and administered by a particular institution. To support defensible assessment practices, university staff might require training and access to assessment and evaluation specialists.

## **Assessing teaching: Methods and metrics**

As discussed previously, there are no universal standards for effective teaching. Thus, many institutions have moved away from defining effective teaching and instead have focused on how it can be measured (Pitman, 2014). For example, in the UK's Teaching Excellence Framework, the Department for Business, Innovation and Skills (BIS; 2015) it is stated that "there is no commonly agreed definition of what constitutes good teaching in higher education .... In the absence of any agreed definition or recognized measures of teaching quality, the Government is proposing to use measures, or metrics, as proxies for teaching quality. Therefore, the challenge is to identify those metrics which most reliably and accurately measure teaching quality, as opposed to other factors that contribute to the results achieved by students" (pp. 5–6). This in and of itself is not necessarily problematic; however, as argued by Saroyan and Amundsen: "it is not so much a problem of whether there are such things as effective teaching behaviours as it is casting a partial picture of teaching in the attempt to assess it" (p. 343, 2001). In other words, the validity of interpretations about teaching are questionable if the assessment is not aligned with effective teaching.

A second and related issue is that evidence of reliability and validity for assessments of teaching are rarely empirically established. As summarized by Strang et al. (2016) in a systematic review that explores defining and assessing quality university teaching in the United Kingdom, there is little evidence contained in this literature that links typically used indicators of effective teaching to positive student outcomes. Although evidence-based recommendations for the methods and metrics of assessing teaching are currently limited, there are a number of common assessment methods that are being used internationally. In the sections below, we briefly review these methods and discuss their strengths and limitations.

## **Student evaluations of teaching**

Student course ratings, also referred to as teaching evaluations or student surveys, are by far the most commonly used method for assessing teaching and are often used to make

---

personnel decisions, such as hiring, tenure, promotions and salary increases (Clayson, 2013; Clayson and Haley, 2011; Jones, 2012; Stowell et al., 2012). This method consists of items that students respond to using quantitative or qualitative input after receiving instruction, typically near the end of a course. The advantage of this method is that it is both cost effective and an efficient method for measuring items that are used as indicators of effective teaching. Student perceptions of their learning experience can provide valuable data and an important outlet for student voice. Another advantage is that this method can provide a more reliable and valid representation of teaching quality than any other method of evaluating teaching (Berk, 2005, 2013; McKeachie, 1979).

While this method has several advantages, there are also a number of justifiable concerns. Open questions remain regarding the relative reliability of data and the validity of interpretations drawn from these measures. Some student evaluations of teaching are rigorously developed according to measurement and evaluation standards while others seem to be locally designed according to “common sense” (Coffey and Gibbs, 2001). This means that the relative quality of student evaluation instruments is variable. The inconsistencies in measurement quality have incited faculty concerns in using this method for accountability purposes (Simpson and Siguaw, 2000; Stein et al., 2012; Surgenor, 2013). One circumstance that might have led to questionable instrument quality is that student evaluations of teaching were not originally designed for making personnel decisions (Hobson and Talbot, 2001). The emerging use of student evaluations of teaching for accountability purposes call for careful evaluations of this method.

Evidence suggests that the results of student evaluations of teaching in higher education can lead to incorrect interpretations of the quality of teaching. Several studies conducted in a variety of contexts suggest that students’ perceptions of the quality of teaching can be influenced by a number of factors such as the gender of the student, the gender of the teacher, the domain of study, type of institution, year of study, the attractiveness of the teacher and prior beliefs about the teacher (Freng and Webber, 2009; Hammonds et al., 2017; Lewandowski et al., 2011; MacNeill et al., 2015; Nargundkar and Shrikhande, 2014; Nasser-Abu, 2017; Onwuegbuzie et al., 2007; Zhao and Gallant, 2012). In other words, factors that are unrelated to the actual quality of teaching seem to systematically influence results from these assessments. This calls into question the reliability and validity of this method. In addition, findings pertaining to individual characteristics of the teacher (for example, gender) also raise concerns regarding the fairness of this assessment practice as it seems to favour certain teaching staff over others. On the other hand, the literature also provides evidence to suggest that factors unrelated to teaching do not significantly influence student evaluations of teaching (d’Appolonia and Abrami, 1997; Remedios and Lieberman, 2008; Marsh, 2007; Marsh and Roche, 1997). These inconsistent findings from the literature call for careful local examination of student evaluations of teaching in accordance with measurement and evaluations standards.

The literature also presents inconsistent findings regarding associations between student evaluations of teaching and student learning. It is expected that effective teaching should relate to student learning. However, empirical evidence for this link is inconsistent. In a recent meta-analysis of 51 studies that examined the association between student evaluations of teaching and student learning, the authors concluded that students do not learn more from teachers who received higher student evaluations of teaching ratings (Uttl et al., 2017). This finding provides compelling evidence that student evaluations of teaching are not associated with student learning. It is worth noting, however, that other individual studies suggest that low to moderate correlations exist between student evaluations of teaching and grades or expected grades (Abrami, 2001; Abrami et al., 1980; Eiszler, 2002; Feldman, 1976; Greenwald and Gillmore, 1997; Stumpf and Freedman, 1979). These inconsistent findings call into question the validity of this assessment method as a true measure of effective teaching. In fact, student evaluations of teaching are largely limited to the delivery of instruction, meaning that other aspects of teaching, including planning and reflection, are

---

easily captured by this method (Saroyan and Amundsen, 2001). It is possible that the partial picture of teaching measured by student evaluations can account for these inconsistent and, at times, surprising empirical results.

Another concern is that the use of student evaluations of teaching can have unintended consequences for teaching and learning. It has been argued that items included in student evaluations might be inconsistent with institutional and pedagogical goals. For example, some have argued that certain student evaluation instruments might favour teacher-centered approaches rather than student-centered approaches to teaching (Abrami et al., 2007). As a result, “it would be foolhardy for young instructors hoping for tenure to create active-engagement learning environments that depend more on students working collaboratively, with or without technology enhancements, despite the overwhelming evidence that such active learning strategies improve learning, knowledge retention and persistence in science studies” (Abrami et al., 2007, p. 451). There is also evidence to suggest that student evaluations of teaching can impede pedagogical innovations in the classroom as trying new methods, which often require trial, error and refinement, can lead to poorer evaluations of teaching (Darwin, 2017). It has also been proposed that teachers might be encouraged to adopt a less challenging curriculum to mitigate poor evaluations from students (Darwin, 2017). It is important then that institutions evaluate the potential unintended consequences that student evaluations of teaching can have, especially when they are used for accountability purposes.

Overall the literature does not provide compelling evidence that student evaluations of teaching allow for reliable and accurate appraisals of teaching effectiveness. This conclusion should urge caution when using this method to make judgements about teaching, especially given the weight that these evaluations can carry for personnel decisions. In order to have confidence in assessment results, institutions should analyse assessments for possible issues with reliability, validity and fairness, and for potential unintended consequences. The findings from such analyses can inform improving the instrument for future administrations and can facilitate appropriate and fair interpretations of assessment results. It also may be appropriate to disseminate the analysis strategy and findings to relevant stakeholders including faculty to improve transparency and to guide the interpretation of results (Winer et al., 2012). In addition, intuitions should be aware that student evaluations of teaching can only capture what students are able to observe in the classroom which means that other important aspects of effective teaching are not measured.

## **Peer observations of teaching**

Peer observations of teaching involve inviting an individual into the classroom to observe teaching in practice. These observations provide an additional lens through which teaching effectiveness can be assessed. It has been suggested that some aspects of quality teaching are best observed by peers, including course design which is not easily recognized by students (Paulsen, 2002). However, assessments of effective teaching using this method are still limited by what is overtly observable in the classroom. These observations are typically structured and observations are guided by an assessment instrument to direct attention to specific aspects of teaching practice. Peer observations can be used for improvement and accountability purposes. When used for improvement purposes, it has been suggested that peer observations can assist teachers to become more self-aware, confident, passionate and willing to try new pedagogical approaches (Cosh, 1998). However, concerns have been raised regarding the use of this assessment method for accountability purposes. At this time, there is little empirical evidence regarding the reliability and validity of this assessment method and best practices are unclear (Pounder et al., 2016).

There are open questions regarding the selection of the observer. Depending on the model, the peer observer can be another faculty member (within or external to the

---

department), a pedagogical expert or a trained student. One drawback of conducting peer observations with teachers from within the same department is that there might be unbalanced power dynamics (McMahon et al., 2007) and can shift the attention away from teaching practice to the content being taught (Torres et al., 2017). Peer observation conducted between different disciplines might be advantageous because it can reduce the focus on the content being taught to the actual practice of teaching (Kinchin, 2005). There are also concerns that “faculty as peer observers, regardless of discipline, are ill-equipped, without further training, to evaluate and provide feedback on the effectiveness of others’ teaching” (Gosling 2009, p. 7). In a similar vein, few academics are educational specialists and therefore might not be qualified to evaluate effective teaching (Smith, 2012).

The major concern with peer observations of teaching is bias. This is particularly problematic when this method is used for accountability purposes. There are many types of bias that are well documented in observation research that could influence the reliability and validity of results from this approach to assessing teaching. One type of potential bias occurs when the observer “has a preconceived expectation of the behaviour of the observed person, s/he is likely to align the actual observations to these expectations to at least some extent” (Muijs, 2006, p. 63). Observer bias has been shown to influence judgements about the effectiveness of teaching. For example, in one experimental study students were told in advance that a visiting lecturer had either a warm or a cold personality. Findings of this study showed that students who were told that the instructor was warm rated him/her as more effective than students who were told he/she was cold (Widmeyer and Loy, 1988). These preconceived expectations among faculty might be difficult to avoid and the amount of bias might be difficult to quantify.

There are a number of strategies that can be implemented to improve the reliability of this method. For example, conducting multiple observations over time and using more than one observer can improve the accuracy of assessment results (Braskamp and Ory, 1994; Centra, 1993; Chism, 1999; DeZure, 1999; Muijs, 2006). In addition, it is recommended that observers undergo specialized training. Although it is unclear what type of training is needed so that peer observers can accurately report on the effectiveness of teaching. For example, Muijs (2006) recommends that “this training needs to familiarize [the observer] with the instrument used, take them carefully through all the steps they need to follow, explain how the data will be interpreted and explain the possible sources of bias they need to be aware of in order not to succumb to them. A significant amount of practice is needed, which should continue until a suitably high level of reliability has been reached” (p. 64). There is some evidence that such training can be successful. For example, in one case study of a programme of peer observation of teaching in Hong Kong, China, students were trained as “consultants” to provide feedback on teaching based on regular non-judgmental observations from the students’ perspective (Pounder et al., 2016).

In summary, peer observations of teaching can measure aspects of effective teaching not easily captured by other methods. Nonetheless, open issues remain regarding efficiency and best practices to reduce bias and improve reliability. Ideally peer observations of teaching should be conducted several times and by multiple trained observers. Given the limited empirical support available to provide evidence for the reliability and validity of this method, institutions should conduct internal evaluations of this method to ensure high-quality assessments. These issues are particularly important when results will be used for accountability purposes.

## Teaching portfolios

Teaching portfolios are increasingly accepted as a comprehensive method of providing evidence of effective teaching in higher education. Portfolios are created by the instructor and contain a range of evidence to demonstrate the quality of instruction based on

---

predetermined guidelines. Evidence is usually provided through artefacts or documentation and can include course outlines, self-reflections, self-evaluations of teaching, a statement of teaching responsibilities, samples of student work and so on. This method can capture aspects of effective teaching that are not easily observable in the classroom. It has also been suggested that this method is the most promising for improvement purposes as portfolios give teachers the opportunity to reflect on their teaching practice (Saroyan and Amundsen, 2001).

When used for accountability purposes, there are concerns regarding the reliability and validity of this method. For example, the evidence provided in a teaching portfolio might be biased if only a partial and favourable picture of teaching quality is provided. Guidelines for constructing teaching portfolios must ensure that the evidence provided is representative of teaching quality and student learning (Abrami et al., 2007). Representativeness is important because “submitting the best student products as evidence of teaching effectiveness, a common practice, does little to allow accurate judgements of how well instructors promote student learning” (Abrami et al., 2007, p. 393). Another open question pertains to how teaching portfolios can be objectively evaluated. The reliability and validity of this method can be improved by establishing assessment rubrics and standards, providing portfolio design guidelines, offering training to reviewers and assigning multiple reviewers (Seldin, 1993; Paulsen, 2002). With regard to training, for example, it has been recommended that reviewers be provided “opportunities to discuss methods, criteria and standards for assessment using portfolios that have previously been rated high or low” (Paulsen, 2002, p. 13).

## **Student learning outcomes**

Another indicator of effective instruction is students’ fulfilment of learning outcomes. Learning outcomes define what students should know and do after instruction and can include the acquisition of knowledge, skills and attitudes. Effective teaching is an important factor for students’ attainment of learning outcomes (Biggs, 2001). Learning outcomes are a broader extension of the use of competencies stipulated by professional organizations. The emphasis on learning outcomes in higher education has become more important in recent years to provide a measure of how much students learn (Tremblay et al., 2012). Typically, learning outcomes are established locally within courses and programmes; although more recently, calls have been made for international and consistent measures of learning outcomes. This notion has been promoted by the OECD through the pilot project entitled “Assessment of Higher Education Learning Outcomes (AHELO)” (Tremblay et al., 2012). As part of this project, the OECD proposed a set of complementary learning outcomes at the global level to guide institutions of higher education to benchmark student performance (Tremblay et al., 2012). Measures of student learning outcomes might provide a promising indicator of effective teaching in higher education; however, some authors have urged that some caution is warranted for several reasons.

Some authors have noted that relative attainment of learning outcomes is caused by a combination of several factors including “teaching, practice and behavioural skills, and of other components that are external to the institution’s capacity” (Hénard, 2010, p. 84). Students’ relative attainment of learning outcomes can therefore be a very noisy and unreliable indicator of effective teaching. The relative contribution of teaching in student attainment must be isolated from these other factors to make judgements about the quality of instruction. Therefore, raw attainment of learning outcomes might not be a reliable indicator of effective teaching rendering the validity of interpretations from this measure questionable. Others have also noted that using students’ attainment of learning outcomes as indicators of effective teaching is problematic because the reliability and validity of assessment practices used to infer student performance within courses might be questionable (Lodge and Bonsanquet, 2014; Sadler, 2009). If the course-level assessments used to

---

measure student performance on learning outcomes are not reliable, valid or fair, then, by extension, interpretations made from these assessment results will also be questionable. Some authors have gone as far as to say that the only way to determine ways in which teaching leads to student learning is through carefully designed, randomized experiments (Bowman and Seifert, 2011).

Advanced statistical methods might be able to address some of these concerns. Over the last decade value-added models have been gaining popularity in K-12 education (for example, OECD, 2008; National Research Council, 2010). Recently, the utility of these models has been considered in higher education for the purpose of comparing the institutional effectiveness for supporting student learning (for example, Cunha and Miller, 2014; OECD, 2013). These models can be used to provide unbiased estimates of effective teaching by tracking academic growth across several years. In the context of K-12 education, some have argued that value-added models have “helped move the conversation about teacher quality to where it belongs – on increasing student learning as the primary goal of teaching. It has also introduced the promise of a much-needed quantitative component in teaching evaluation, while prompting a re-examination of issues of fairness and proper test use” (Braun, 2005, p. 9). A growing body of experimental and quasi-experimental research in K-12 education has provided some initial support that these modelling techniques can estimate the impact of teaching on student test score growth (Chetty et al., 2014; Kane and Staiger, 2008; Kane et al., 2013). Despite this evidence, in practice there remains the fundamental concern that “if making casual attributions is the goal, then no statistical model, however complex, and no method of analysis, however sophisticated, can fully compensate for the lack of randomization” (Braun, 2005, p. 8). More research “is needed to understand how [value added model] estimates should (or should not) be combined with other metrics to identify and retain effective teachers” (Chetty et al., 2014). The lessons learned from K-12 education can provide important insights into the use of this method for assessing teaching in higher education that are worth exploring.

This review of current methods and metrics for assessing teaching in higher education suggests that no single method can measure effective teaching. Rather, it is generally agreed that in order to assess and measure effective teaching in higher education multiple assessment methods are required (Hammonds et al., 2017), including new and innovative methods (Abrami et al., 2007; Hénard, 2010; Saroyan and Amundsen, 2001). This can include triangulating multiple types of assessment evidence to evaluate teaching (Hénard, 2010; Hénard and Roseveare, 2012; Lodge and Bonsanquet, 2014; Saroyan and Amundsen, 2001; Spooen et al., 2013). Aligning assessment evidence from different sources can provide a more holistic and accurate picture of teaching that no single method can capture. Selecting these methods strategically can also have the benefit of leveraging the strengths of particular methods to mitigate the weakness of others. Fundamentally, assessment methods should be evaluated for their reliability, validity and fairness so that institutions can place the appropriate weight on assessment results (Uttl et al., 2017). For example, results from assessments methods that have been systematically evaluated and are deemed reliable, valid and fair should be weighted more than methods with uncertain results.

Specialized training for staff might be necessary to support the design, implementation, analysis and interpretation of assessment results (Hénard and Roseveare, 2012; Linse, 2017; Spooen et al., 2013). For example, training for staff who administer and interpret assessments of teaching could include both qualitative and quantitative methods and educational theory. In addition, it may be beneficial to provide teachers with access to measurement and evaluation specialists who can assist them to interpret and understand the results of their teaching evaluations.

---

## Emerging trends and issues

In this final section, we explore two emerging trends and issues for assessing teaching in higher education. The first pertains to the casualization of teaching staff and the second considers the unique context of e-learning and online teaching. We discuss the specific considerations and implications of these new trends for assessing teaching.

### Casualization of teaching staff

The massification of higher education has created new staffing demands for institutions of higher education. This has led to a global increase in the casualization of teaching staff, including adjuncts, casuals, contract, contingent and non-tenure track faculty (Bryson, 2013; Coates and Goedegebuure, 2010; Parker, 2012; Harvey, 2017). These staff are typically employed on an hourly, contract or honorary basis (Percy et al., 2008). In Australia, for example, the majority of university teaching is now complemented by casual teachers (May et al., 2013). In the United States, there is also a reliance on non-tenure tract faculty for teaching (Jaschik and Lederman, 2018). The number of casual teaching staff is also on the rise in Canada, Japan, New Zealand and Europe (Bryson, 2013; Sutherland and Gilbert, 2013; Sursock, 2015). This trend is predicted to continue or even increase (Jaschik and Lederman, 2018) as the employment of casual teachers is generally more cost effective and can “free up” faculty for research (Bryson, 2013). For example, in a social science department at one research-intensive university in the United Kingdom, casual teaching staff were responsible for all undergraduate courses (Powney et al., 2003). It has been suggested that the increase in casual teaching staff has not been met with an escalation in approaches to ensure the quality of learning and teaching (Harvey, 2017).

Data on casual teaching staff is limited due to the constantly changing employment of these individuals (Bryson, 2013). This “makes it more challenging to develop systematic approaches to recognizing and supporting these staff if we do not know who they are, where they are located and what professional development needs they have” (Harvey, 2017, p. 1). Thus, the casualization of teaching in institutions of higher education creates new challenges for methodically assessing, tracking and supporting the quality of teaching completed by casual staff. A salient issue is the inequality of opportunity for casual teaching staff with regard to support and treatment (Bryson, 2013). For example, these staff members often do not have access to professional development for induction, mentoring, course design and training (Anderson, 2007). Others have noted that casual teaching staff often have difficulty accessing resources that are essential to their teaching, including library services and office space (Kimber, 2003). This environment might create inequalities in the opportunity to perform. Specifically, the lack of access to resources can create barriers for casual teaching staff to engage in effective teaching. This can have important implications for staff retention and reappointment. Some have argued that the pressure to perform well on student evaluations of teaching without the necessary resources can have unintended negative consequences for the quality of student learning. For example, several studies have suggested that casual teaching staff typically give higher grades than their full-time colleagues (for example, Fedler et al., 1989; Goldberg and Callahan, 1991; Salamonson et al., 2010). One of the possible reasons for this discrepancy is that casual teaching staff can be more lenient in order to secure better student evaluations of their teaching (Moore and Trahan, 1998; Salamonson et al., 2010). Therefore, the lack of requisite training and resources in combination with the pressure to perform might have unintended consequences for the quality of teaching and learning. Therefore, a closer look at how these staff members are supported and assessed is warranted. Some have suggested that supporting “comprehensive and accurate data collection”, including the effectiveness of teaching, can be used to signal professional development needs (Harvey, 2017).

---

## E-learning and teaching online

A second emerging trend is the increase in online courses and distance education. Over the last decade, the number of students taking online courses has grown by millions (Allen and Seaman, 2015). Many institutions of higher education are including online learning as part of their long-term strategy (Allen and Seaman, 2011; He et al., 2014). The proportion of content delivered online can vary based on the type of course. As defined by Allen and Seaman (2015), web-facilitated courses deliver 1–29 per cent of course material online and they are typically delivered in face-to-face learning environments but use learning management systems or web pages to post the syllabus and assignments. Blended or hybrid courses use a combination of online and face-to-face instruction, where a substantial amount of the content is supported online (30–79 per cent), including, for example, online discussions, and has fewer face-to-face meetings. Online courses typically have no face-to-face meetings with all of the content delivered online.

As this is an emerging area in the scholarship of teaching and learning, there are many open questions that require further investigation. Three such issues are briefly discussed. The first issue pertains to the operationalization of effective online teaching. It is currently unclear how to define effective teaching in online environments and how teaching online might differ from face-to-face learning environments. This first issue has two implications for assessing teaching in higher education. The first is that students might value online teaching in different ways than they value teaching in traditional learning environments (Young, 2006, p. 65). The second is that there are unique factors required to ensure effective teaching in online contexts (Sun et al., 2008). For example, it has been suggested that effective teaching in e-learning environments requires technical skills, online moderation skills and online communication skills (Gunn, 2001). This means that current assessments of teaching need to be adjusted or redesigned to measure relevant indicators that will allow for valid interpretations of effective teaching in these environments. This is because “designing and implementing strategies to evaluate online teaching necessarily differs from designing and implementing those used in the face-to-face environment; they vary as a function of the nature, purpose and focus of the evaluation” (Tobin et al., 2015, p. 15). For example, student evaluations of teaching instruments designed for face-to-face learning environments do not include items that can assess quality teaching in online instruction (Bangert, 2006). It is also unclear as to how other methods for assessing teaching used in face-to-face environments can be adapted for online environments. It is unclear how observations of teaching can be utilized in online teaching situations. Therefore, attempts to apply existing assessment methods might not be appropriate for online environments. A second issue is that it is difficult to separate teaching from the design elements of the learning management system (Tobin et al., 2015, p. 15). Assessments of teaching in online environments should be designed to measure only dimensions that are within the teacher’s control.

There are also new opportunities for assessing teaching in online environments that are not possible in face-to-face learning environments. Learning management systems can collect instructor trace data that could be used as indicators of effective teaching. Indicators worth considering include frequency of logins, time spent on learning management systems, nature and quality of feedback, timeliness of grading and feedback and proportion of instructor-to-student posts (Tobin et al., 2015, p. 15). While this new opportunity exists, evidence to justify the use of such indicators is still needed. Even though learning management systems can provide new sources of data on instructor interactions, it is important to determine which of these are relevant for making inferences about effective teaching and which are not.

---

## Summary of recommendations

Reviewing current measurement and evaluation standards and methods and metrics of assessing teaching in higher education suggests several policy recommendations. These are summarized below:

- The trend of quality assurance in higher education has created new challenges for institutions to develop and demonstrate effective teaching. However, conceptualizations of effectiveness are highly contextualized and measurements of teaching are largely inconsistent between institutions, contexts, disciplines and jurisdictions. Policies could be developed to facilitate the adoption of consistent standards of effective teaching and quality assessment practices, which can include promoting the use of multiple assessment methods to capture the multidimensionality of effective teaching.
- Quality assurance also calls for policies that can guide the appropriate use of measurement and evaluation principles in assessments of teaching. Policies could stipulate how to apply the principles of reliability, validity and fairness for developing, implementing and interpreting the results from assessments of teaching. Policies can also guide standards for documenting evidence to support claims that assessment practices are reliable and allow for valid and fair interpretations of assessment results. Such claims are necessary to have confidence in interpretations about teaching, which have important ramifications for personnel decisions and for monitoring and maintaining quality education. Emerging trends should also be considered including for developing casual teaching staff and for online teaching contexts.
- Relatedly, guidance is needed for how institutions can respond if assessment of teaching is found to be unreliable or leads to inaccurate or unfair interpretations about teaching. For example, if a student evaluation of teaching instruments has been found to systematically and negatively bias results of, for example, female instructors, the results from this assessment should be statistically corrected or appropriately weighted and the assessment instrument should be adjusted for future use.
- Policies should be developed and communicated to teaching staff to encourage transparency in defined standards of effective teaching, how it is measured and how results are used within institutions. For example, institutions could be encouraged to develop and provide internal assessment and evaluation policies. Transparent internal policies can help ensure consistent assessment practices within institutions and can contribute to monitoring and maintaining quality education. Such policies can also communicate to teachers what is expected of them to be recognized as an effective teacher.
- There should also be alignment between institutional policies to indicate that teaching is valued, the choice of indicators of effective teaching, assessment tools and incentives.
- Policies should encourage institutions to provide opportunities for training and professional development to facilitate quality assessment practices. Training will become increasingly important should institutions move towards more sophisticated statistical methods for assessing teaching, such as value-added models. Training should be provided for staff in charge of developing, administering and interpreting assessment results. For example, staff should receive training in educational measurement and evaluation theories, assessment methodologies, understanding of assessment development, administration and interpretation including principles of reliability, validity and fairness and qualitative and quantitative analyses. In addition, it may be beneficial to provide teachers with access to measurement and evaluation specialists

---

who can assist them to interpret and understand the results of their teaching assessments. In addition, credible specialists should be available within institutions to provide professional development of teaching staff with respect to teaching.

## Areas for future research

International data on assessing teaching in higher education for development and accountability purposes is lacking or difficult to obtain. To provide evidence-based recommendations of best practices, additional research in several areas is needed. These areas for future research are summarized below:

- The development of internationally accepted frameworks or standards for defining and assessing effective teaching: these standards can help ensure the quality of higher education globally. However, given the complex nature of teaching in higher education, standards for teaching in higher education require further research.
- A review of how institutions collect, weight and use different methods of assessing teaching in higher education: the aim would be to provide initial benchmarking of assessment practices across institutions and could include internal policies, methods and metrics, evidence of assessment quality, resources and training and uses.
- Further exploration of emerging assessment methods including value-added models and trace data of online teaching within the context of higher education: this research would help clarify the utility of these methods and could inform the development of new guidelines and policies.
- Compilation of international data on casual teaching staff, including their professional development needs.

## Conclusions

The massification of higher education, increase in quality assurance, the changing composition of teaching staff, growth in online teaching and a greater need for professional development, have created new demands and challenges for assessing teaching in higher education. More importantly, they have brought to the fore the necessity of quality assessment systems that enable institutions to evaluate the effectiveness of teaching for improvement and accountability purposes. This brief review has highlighted the complexity of developing such systems as well as opportunities for advancing assessment practices in higher education. Assessment of teaching must be grounded in accurate conceptualizations of effective teaching to set standards and expectations for teaching. At this time, global standards of effective teaching have yet to be established but this goal can be achieved by building on locally defined definitions. In addition, the development, administration and interpretation of assessment of teaching must adhere to recognized standards of measurement and evaluation. This means that staff responsible for assessing teaching must have access to training opportunities to ensure the quality of institutional assessment systems. Although this review was limited by the availability of international data on assessing teaching in higher education, it provides a good foundation of information to determine the current state of assessing teaching in higher education and where gaps exist. To ensure quality education in the higher education sector, it is essential that quality evidence-based systems for assessing teaching are supported.



---

## References

- Abrami, P.C. 2001. "Improving judgments about teaching effectiveness using teacher rating forms", in *New Directions for Institutional Research*, 27(5), pp. 59–87.
- Abrami, P.C.; d'Apollonia, S.; Rosenfield, S. 2007. "The dimensionality of student ratings of instruction: What we know and what we do not", in *The scholarship of teaching and learning in higher education: An evidence-based perspective*, pp. 385–456: Springer.
- Abrami, P.C.; Dickens, W.J.; Perry, R.P.; Leventhal, L. 1980. "Do teacher standards for assigning grades affect student evaluations of instruction?", in *Journal of Educational Psychology*, 72, 1, pp. 107–118.
- Allen, I.E.; Seaman, J. 2011. *Going the distance: Online education in the United States*, 2011. Sloan Consortium, PO Box 1238, Newburyport, MA 01950.
- . 2015. *Grade Level: Tracking Online Education in the United States*, Babson Survey Research Group.
- Altbach, P.; Reisberg, L.; Rumbley, L. 2009. *Trends in Global Higher Education: Tracking an Academic Revolution*, report prepared for the UNESCO 2009 World Conference on Higher Education, UNESCO, Paris.
- American Educational Research Association, American Psychological Association and National Council on Measurement in Education. 2014. *Standards for educational and psychological testing*.
- Anderson, V. 2007. "Contingent and marginalised? Academic development and part-time teachers", in *International Journal for Academic Development*, 12(2), pp. 111–121.
- Bangert, A.W. 2006. "The development of an instrument for assessing online teaching effectiveness", in *Journal of Educational Computing Research*, 35(3), p. 227.
- Berk, R.A. 2005. "Survey of 12 strategies to measure teaching effectiveness", in *International Journal of Teaching and Learning in Higher Education*, 17(1), pp. 48–62.
- . 2013. *Top 10 flashpoints in student ratings and the evaluation of teaching: What faculty and administrators must know to protect themselves in employment decisions*. Stylus Publishing, LLC.
- Biggs, J. 2001. "The reflective institution: Assuring and enhancing the quality of teaching and learning", in *Higher education*, 41(3), pp. 221–238.
- Bowman, N.A.; Seifert, T.A. 2011. "Can college students accurately assess what affects their learning and development?", in *Journal of College Student Development*, 52(3), pp. 270–290.
- Braskamp, L.A.; Ory, J.C. 1994. *Assessing Faculty Work: Enhancing Individual and Institutional Performance*, Jossey-Bass Higher and Adult Education Series, San Francisco, CA: Jossey-Bass Inc.
- Braun, H.I. 2005. "Using student progress to evaluate teachers: A primer on value-added models. Policy Information Perspective", in *Educational Testing Service*.

- 
- Bryson, C. 2013. *Supporting sessional teaching staff in the UK - To what extent is there real progress?*, research online.
- Centra, J.A. 1993. *Reflective Faculty Evaluation: Enhancing Teaching and Determining Faculty Effectiveness*, The Jossey-Bass Higher and Adult Education Series, San Francisco, CA: Jossey-Bass Inc.
- Chetty, R.; Friedman, J.N.; Rockoff, J.E. 2014. "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates", in *American Economic Review*, 104(9), pp. 2593–2632.
- Chetty, R.; Friedman, J.; Rockoff, J. 2014. "Discussion of the American statistical association's statement (2014) on using value-added models for educational assessment", in *Statistics and Public Policy*, 1(1), pp. 111–113.
- Chickering, A.W.; Gamson, Z.F. 1987. "Seven principles for good practice in undergraduate education", in *American Association for Higher Education Bulletin*, pp. 3–7.
- Chism, N.V.N. 1999. *Peer Review of Teaching. A Sourcebook*, Bolton, MA: Anker.
- Chism, N.V.N.; Chism, G.W. 2007. *Peer review of teaching: A sourcebook*, Bolton, MA: Anker Pub. Co.
- Clayson, D.E. 2013. "Initial impressions and the student evaluation of teaching", in *Journal of Education for Business*, 88(1), pp. 26–35.
- Clayson, D.E.; Haley, D.A. 2011. "Are students telling us the truth? A critical look at the student evaluation of teaching", in *Marketing Education Review*, 21(2), pp. 101–112.
- Coates, H.; Goedegebuure, L. 2010. *The real academic revolution*, research briefing, Carlton: LH Martin Institute.
- Coffey, M.; Gibbs, G. 2001. "The evaluation of the student evaluation of educational quality questionnaire (SEEQ) in UK higher education", in *Assessment & Evaluation in Higher Education*, 26, 1, pp. 89–93.
- Cosh, J. 1998. "Peer observation in higher education – A reflective approach", in *Innovations in Education and Training International*, 35, 2, pp. 171–176.
- Cunha, J.M.; Miller, T. 2014. "Measuring value-added in higher education: Possibilities and limitations in the use of administrative data", in *Economics of Education Review*, 42, pp. 64–77.
- d'Apollonia, S.; Abrami, P.C. 1997. "Navigating student ratings of instruction", in *American Psychologist*, 52, 11, pp. 1198–1208.
- Darwin, S. 2017. "What contemporary work are student ratings actually doing in higher education?", in *Studies in Educational Evaluation*, 54, pp. 13–21.
- Department for Business, Innovation and Skills. 2015. *Fulfilling our potential: Teaching excellence, social mobility and student choice*, Department for Business, Innovation and Skills, London: BIS.
- DeZure, D. 1999. "Evaluating teaching through peer classroom observation", in P. Seldin (ed.): *Changing Practices in Evaluating Teaching: A Practical Guide to Improved Faculty Performance and Promotion/Tenure Decisions*, Bolton, MA: Anker.

- 
- Eiszler, C.F. 2002. "College students' evaluations of teaching and grade inflation", in *Research in Higher Education*, 43(4), pp. 483–501.
- Fedler, F.; Counts, T.; Stoner, K.R. 1989. "Adjunct profs grade higher than faculty at three schools", in *Journalism Educator*, 44, pp. 32–37.
- Feldman, K.A. 1976. "Grades and college students' evaluations of their courses and teachers", in *Research in Higher Education: Journal of the Association for Institutional Research*, 4, 1, pp. 69–111.
- Freng, S.; Webber, D. 2009. "Turning up the heat on online teaching evaluations: Does 'hotness' matter"?, in *Teaching of Psychology*, 36, 3, pp. 189–193.
- Goldberg, G.; Callahan, J. 1991. "Objectivity of student evaluations of instructors", in *Journal of Education for Business*, 66, pp. 377–378.
- Gosling, D. 2009. "A new approach to peer review of teaching", in D. Gosling and K. Mason O'Connor (eds): *Beyond the peer observation of teaching*, pp. 7–15. SEDA Paper 124, London: Staff and Educational Development Association Ltd.
- Greenwald, A.G.; Gillmore, G.M. 1997. "Grading leniency is a removable contaminant of student ratings", in *American Psychologist*, 52(11), p. 1209.
- Gunn, C. 2001. "Effective online teaching: How far do the frameworks go", in *Meeting at the Crossroads: Proceedings of the Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCLITE)*, pp. 235–244.
- Hammonds, F.; Mariano, G.J.; Ammons, G.; Chambers, S. 2017. "Student evaluations of teaching: Improving teaching quality in higher education", in *Perspectives: Policy and Practice in Higher Education*, 21(1), p. 26.
- Harvey, L. 2018. [Analytic Quality Glossary](#), Quality Research International.
- Harvey, M. 2017. *Quality learning and teaching with sessional staff: Systematising good practice for academic development*: Taylor and Francis.
- He, W.; Xu, G.; Kruck, S. 2014. "Online IS education for the 21st century", in *Journal of Information Systems Education*, 25(2), p. 101.
- Hénard, F. 2010. *Learning our lesson: Review of quality teaching in higher education* (Vol. 2010, No. 2), Paris, OECD.
- Hénard, F.; Roseveare, D. 2012. "Fostering quality teaching in higher education: Policies and Practices", in *An IMHE Guide for Higher Education Institutions*, pp. 7–11.
- Hobson, S.M.; Talbot, D.M. 2001. "Understanding student evaluations: What all faculty should know", in *College teaching*, 49(1), pp. 26–31.
- Jaschik, S.; Lederman, D. 2018. [The 2018 Inside Higher Ed Survey of College and University Chief Academic Officers](#). Inside Higher Ed.
- Jones, S.J. 2012. "Reading between the lines of online course evaluations: Identifiable actions that improve student perceptions of teaching effectiveness and course value", in *Journal of Asynchronous Learning Networks*, 16(1), pp. 49–58.
- Kane, T.J.; Staiger, D.O. 2008. *Estimating teacher impacts on student achievement: An experimental evaluation*, No. 14607, National Bureau of Economic Research.

- 
- Kane, T.J.; McCaffrey, D.F.; Miller, T.; Staiger, D.O. 2013. "Have we identified effective teachers? Validating measures of effective teaching using random assignment", in *Research Paper*, MET Project, Bill & Melinda Gates Foundation.
- Kember, D.; Ma, R.; McNaught, C. 2006. *Excellent university teaching*, Hong Kong, China: The Chinese University Press.
- Kimber, M. 2003. "The tenured 'core' and the tenuous 'periphery': The casualisation of academic work in Australian universities", in *Journal of Higher Education Policy and Management*, 25(1), pp. 41–50.
- Kinchin, I.M. 2005. "Evolving diversity within a model of peer observation at a UK university", in *British Educational Research Association (BERA) Annual Conference, University of Glamorgan, Wales*, pp. 14–17.
- Lewandowski, G.W. Jr; Higgins, E.; Nardone, N.N. 2011. "Just a harmless website?: An experimental examination of RateMyProfessors.com's effect on student evaluations", in *Assessment & Evaluation in Higher Education*, 37(8), pp. 987–1002.
- Linse, A.R. 2017. "Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees", in *Studies in Educational Evaluation*, 54, pp. 94–106.
- Lodge, J.M.; Bonsanquet, A. 2014. "Evaluating quality learning in higher education: Re-examining the evidence", in *Quality in Higher Education*, 20(1), pp. 3–23.
- MacNell, L.; Driscoll, A.; Hunt, A.N. 2015. "What's in a name: Exposing gender bias in student ratings of teaching", in *Innovative Higher Education*, 40(4), pp. 291–303.
- Marsh, H.W. 2007. "Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness", in *The scholarship of teaching and learning in higher education: An evidence-based perspective*, pp. 319–383: Springer.
- Marsh, H.W.; Roche, L.A. 1997. "Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility", in *American Psychologist*, 52, 11, pp. 1187–1197.
- May, R.; Strachan, G.; Peetz, D. 2013. "Workforce development and renewal in Australian universities and the management of casual academic staff", in *Journal of University Teaching and Learning Practice*, 10(3), p. 3.
- McKeachie, W.J. 1979. "Student ratings of faculty: A reprise", in *Academe*, 65(6), pp. 384–397.
- McMahon, T.; Barrett, T.; O'Neill, G. 2007. "Using observation of teaching to improve quality: Finding your way through the muddle of competing conceptions, confusion of practice and mutually exclusive intentions", in *Teaching in Higher Education*, 12(4), pp. 499–511.
- Moore, M.; Trahan, R. 1998. "Tenure status and grading practice", in *Sociological Perspectives*, 41(4), pp. 775–782.
- Muijs, D. 2006. "Measuring teacher effectiveness: Some methodological reflections", in *Educational research and evaluation*, 12(1), pp. 53–74.

- 
- Nargundkar, S.; Shrikhande, M. 2014. “Norming of student evaluations of instruction: impact of non-instructional factors”, in *Decision Sciences: Journal of Innovative Education*, 12, 1, pp. 55–72.
- Nasser-Abu Alhija, F. 2017. “Teaching in higher education: Good teaching through students’ lens”, in *Studies in Educational Evaluation*, 54, pp. 4–12.
- National Research Council. 2010. *Getting value out of value-added*, H. Braun, N. Chudowsky and J. Koenig (eds) (Washington, DC, National Academies Press).
- Onwuegbuzie, A.J.; Witcher, A.E.; Collins, K.M.T.; Filer, J.D.; Wiedmaier, C.D.; Moore, C.W. 2007. “Students perceptions of characteristics of effective college teachers: A validity study of a teaching evaluation form using a mixed-methods analysis”, in *American Educational Research Journal*, 44(1), pp. 113–160.
- Organisation for Economic Co-operation and Development (OECD). 2008. *Measuring improvements in learning outcomes: Best practices to assess the value-added of schools* (Paris, OECD).
- . 2009. *Teacher Evaluation: A Conceptual Framework and Examples of Country Practices*, Paris.
- . 2013. *Assessment of Higher Education Learning Outcomes: Feasibility Study Report, Vol.3: Further Insights* (Paris, OECD).
- . 2017. *Enhancing Higher Education System Performance*, Paris.
- Parker, L.D. 2012. “From privatised to hybrid corporatised higher education: A global financial management discourse”, in *Financial Accountability and Management*, 28(3), pp. 247–268.
- Paulsen, M.B. 2002. “Evaluating teaching performance”, in *New Directions for Institutional Research*, 2002 (114), pp. 5–18.
- Percy, A.; Scoufis, M.; Parry, S.; Goody, A.; Hicks, M.; Macdonald, I.; Martinez, K. et al. 2008. *The RED Report: Recognition – Enhancement – Development: The contribution of sessional teachers to higher education*, Sydney: Australian Learning and Teaching Council.
- Pitman, T. 2014. “Reinterpreting higher education quality in response to policies of mass education: The Australian experience”, in *Quality in Higher Education*, 20(3), pp. 348–363.
- Pounder, J.S.; Ho Hung-lam, E.; Groves, J.M. 2016. “Faculty–student engagement in teaching observation and assessment: A Hong Kong initiative”, in *Assessment & Evaluation in Higher Education*, 41(8), p. 1193.
- Powney, J.; Bryson, C.; Cloonan, M.; Wilson, V. 2003. *The Appointment, Retention and Promotion of Academic Staff within Higher Education Institutions* (Bristol, HEFCE).
- Prosser, M.; Trigwell, K. 1997. “Relations between perceptions of the teaching environment and approaches to teaching”, in *British Journal of Educational Psychology*, 67(1), pp. 25–35.
- Ramsden, P. 2003. “Learning to teach in higher education”, London: Routledge.

- 
- Remedios, R.; Lieberman, D.A. 2008. "I liked your course because you taught me well: The influence of grades, workload, expectations and goals on students' evaluations of teaching", in *British Educational Research Journal*, 34, 1, pp. 91–115.
- Sadler, D.R. 2009. "Indeterminacy in the use of preset criteria for assessment and grading", in *Assessment & Evaluation in Higher Education*, 34, 2, pp. 159–179.
- Salamonson, Y.; Halcomb, E.J.; Andrew, S.; Peters, K.; Jackson, D. 2010. "A comparative study of assessment grading and nursing students' perceptions of quality in sessional and tenured teachers", in *Journal of Nursing Scholarship*, 42(4), pp. 423–429.
- Salmi, J.; Saroyan, A. 2007. "League tables as policy instruments", in *Higher Education Management and Policy*, 19(2), pp. 1–38.
- Santiago, P.; Tremblay, K.; Basri, E.; Arnal, E. 2008. *Tertiary education for the knowledge society*, Vol. 1 (Paris, OECD).
- Saroyan, A.; Amundsen, C. 2001. "Evaluating university teaching: Time to take stock", in *Assessment and Evaluation in Higher Education*, 26(4), pp. 341–353.
- Saroyan, A.; Amundsen, C.; McAlpine, L.; Weston, C.; Winer, L.; Gandell, T. 2004. "Assumptions underlying workshop activities", in A. Saroyan and C. Amundsen (eds): *Rethinking teaching in higher education*, pp. 15–29, Sterling, VA: Stylus.
- Seldin, P. 1993. *Successful use of teaching portfolios*, Bolton, MA: Anker Pub. Co.
- Simpson, P.M.; Siguaw, J.A. 2000. "Student evaluations of teaching: An exploratory study of the faculty response", in *Journal of Marketing Education*, 22(3), pp. 199–213.
- Skelton, A. 2005. *Understanding teaching excellence in higher education: Towards a critical approach*, London: Routledge.
- Smith, H. 2012. "The unintended consequences of grading teaching", in *Teaching in higher education*, 17(6), p. 747.
- Spooren, P.; Brockx, B.; Mortelmans, D. 2013. "On the validity of student evaluation of teaching: The state of the art", in *Review of Educational Research*, 83(4), pp. 598–642.
- Stanfield, D.A.; Shimmi, Y. 2014. "Chinese higher education: Statistics and trends", in L.E. Rumbley, R.M. Helms, P.M. Peterson, P.G. Altbach (eds): *Global Opportunities and Challenges for Higher Education Leaders*, pp. 77–82, Rotterdam: Sense Publishers.
- Stein, S.J.; Spiller, D.; Terry, S.; Harris, T.; Deaker, L.; Kennedy, J. 2012. "Unlocking the impact of tertiary teachers' perceptions of student evaluation of teaching", Wellington, New Zealand: Ako Aotearoa National Centre for Tertiary Teaching Excellence.
- Stowell, J.R.; Addison, W.E.; Smith, J.L. 2012. "Comparison of online and classroom-based student evaluations of instruction, in *Assessment & Evaluation in Higher Education*, 37(4), pp. 465–473.
- Strang, L.; Bélanger, J.; Manville, C.; Meads, C. 2016. *Review of the Research Literature on Defining and Demonstrating Quality Teaching and Impact in Higher Education*, York, United Kingdom: Higher Education Academy.

- 
- Stumpf, S.A.; Freedman, R.D. 1979. "Expected grade covariation with student ratings of instruction: Individual versus class effects", in *Journal of Educational Psychology*, 71(3), pp. 293–302.
- Sun, P.-C.; Tsai, R.J.; Finger, G.; Chen, Y.-Y.; Yeh, D. 2008. "What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction", in *Computers and Education*, 50(4), pp. 1183–1202.
- Surgenor, P.W.G. 2013. "Obstacles and opportunities: Addressing the growing pains of summative student evaluation of teaching", in *Assessment & Evaluation in Higher Education*, 38, pp. 363–376.
- Sursock, A. 2015. "Trends 2015: Learning and teaching in European universities", in *Brussels: European University Association*.
- Sutherland, K.A.; Gilbert, A. 2013. "Academic aspirations amongst sessional tutors in a New Zealand university", in *Journal of University Teaching & Learning Practice*, 10(3), p. 7.
- Tobin, T.J.; Mandernach, B.J.; Taylor, A.H. 2015. *Evaluating online teaching: Implementing best practices*, John Wiley & Sons.
- Torres, A.C.; Lopes, A.; Valente, J.M.S.; Mouraz, A. 2017. "What catches the eye in class observation? Observers' perspectives in a multidisciplinary peer observation of teaching program", in *Teaching in Higher Education*, 22(7), p. 822.
- Tremblay, K.; Lalancette, D.; Roseveare, D. 2012. *Assessment of Higher Education Learning Outcomes: Feasibility Study Report*, Vol. 1, OECD, Paris.
- The World Bank. [The World Bank Education Statistics](#).
- Trigwell, K.; Prosser, M. 1996. "Changing approaches to teaching: A relational perspective", in *Studies in Higher Education*, 21(3), pp. 275–284.
- Uttl, B.; White, C.A.; Gonzalez, D.W. 2017. "Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related", in *Studies in Educational Evaluation*, 54, pp. 22–42.
- Vincent-Lancrin, S. 2008. "What is the impact of demography on higher education systems? A forward-looking approach for OECD countries", in *Higher Education to 2030*, pp. 41–103.
- Widmeyer, W.N.; Loy, J.W. 1988. "When you're hot, you're hot! Warm–cold effects in first impressions of persons and teaching effectiveness", in *Journal of Educational Psychology*, 80(1), p. 118.
- Winer, L.; Di Genova, L.; Vungoc, P.-A.; Talsma, S. 2012. *Interpreting end-of-course evaluation results*, Montreal: Teaching and Learning Services, McGill University.
- Young, S. 2006. "Student views of effective online teaching in higher education", in *American Journal of Distance Education*, 20(2), p. 65.
- Zhao, J.; Gallant, D.J. 2012. "Student evaluation of instruction in higher education: Exploring issues of validity and reliability", in *Assessment & Evaluation in Higher Education*, 37(2), pp. 227–235.