Quality of teaching in higher education

Noben, Ine; Deinum, Jan Folkert; Hofman, W. H. Adriaan

Published in:
International Journal for Academic Development

DOI:
10.1080/1360144X.2020.1830776

International Journal for Academic Development

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rija20

Quality of teaching in higher education: reviewing teaching behaviour through classroom observations

Ine Noben, Jan Folkert Deinum & W. H. Adriaan Hofman

To cite this article: Ine Noben, Jan Folkert Deinum & W. H. Adriaan Hofman (2022) Quality of teaching in higher education: reviewing teaching behaviour through classroom observations, International Journal for Academic Development, 27:1, 31-44, DOI: 10.1080/1360144X.2020.1830776

To link to this article: https://doi.org/10.1080/1360144X.2020.1830776

INTERNATIONAL JOURNAL FOR ACADEMIC DEVELOPMENT
2022, VOL. 27, NO. 1, 31–44
https://doi.org/10.1080/1360144X.2020.1830776

Routledge
Taylor & Francis Group

Quality of teaching in higher education: reviewing teaching behaviour through classroom observations

Ine Noben, Jan Folkert Deinum and W. H. Adriaan Hofman

Teacher Education Department, Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, The Netherlands

ABSTRACT

This study examines the extent to which lecturers demonstrate effective teaching behaviour. The results of 203 observations reveal substantial differences in detected teaching behaviour. Lecturers mostly demonstrated teaching behaviour in the domains classroom climate, efficient organisation, and instruction. Teaching behaviour relating to the domains activating teaching and teaching learning strategies was observed less frequently, with almost no evidence of behaviour associated with the domain differentiation. The quality of teaching in small classes was slightly higher than that in large classes. These findings can help tailor professional learning activities to lecturers' professional development needs.

ARTICLE HISTORY
Received 8 November 2019
Accepted 7 June 2020

KEYWORDS
Classroom observation; quality of teaching; teaching behaviour; university teaching

Introduction

Effective professional learning activities focus on meeting the professional development needs of individual lecturers (Jacob et al., 2015). To balance various needs, educational developers mainly support lecturers in two central practices: course planning and classroom teaching (Hora & Ferrare, 2013). While considerable literature has been published on course planning practices, relatively few studies have focussed on characterizing lecturers' classroom instruction. This is surprising since research has clearly shown the quality of teaching behaviour is associated with student achievement (Hattie, 2015; Schneider & Preckel, 2017). Furthermore, the descriptions of teaching practices in the literature rely mostly on lecturers' self-reports, while actual teaching behaviour is hardly captured (Stes et al., 2010). To fill this gap, our study aims to identify to what extent lecturers demonstrate effective teaching behaviour. Data will be collected through classroom observations. Whereas other researchers have examined the frequency of teaching behaviour in the higher education classroom (Hora & Ferrare, 2013; Stains et al., 2018), this study focuses on the quality of teaching behaviour.

CONTACT Ine Noben i.noben@rug.nl

This article has been corrected with minor changes. These changes do not impact the academic content of the article.
Supplemental data for this article can be accessed at https://doi.org/10.1080/1360144X.2020.1830776.

Theoretical framework

Effective teaching behaviour

The literature has long focussed on dichotomising teaching practices in either teacher-focused or student-focused behaviour (Postareff et al., 2008; Trigwell & Prosser, 2004), where the latter is considered to be more effective. However, teaching is complex and various teaching behaviours affect student achievement (Hattie, 2015; Schneider & Preckel, 2017). The authors have recognised this and have recently moved beyond the dichotomisation of teaching by categorising teaching behaviour in various domains (Hora & Ferrare, 2013; Stains et al., 2018). The categorisation still simplifies teaching to a certain extent, but helps educational developers to support lecturers in developing their teaching behaviour by focusing feedback on specific teaching domains (Maulana et al., 2015). From the literature we derived six observable teaching behaviour domains that are known to impact student achievement: (1) safe and stimulating learning climate, (2) efficient organisation (classroom management), (3) clarity of instruction, (4) intensive and activating teaching, (5) teaching learning strategies, and (6) differentiation.

A safe classroom learning climate has been defined as a climate 'that allows students to feel secure enough to take risks, honestly express their views, and share and explore their knowledge, attitudes, and behaviours' (Holley & Steiner, 2005, p. 50). Teaching behaviour that promotes such a climate has been found to have a medium-large effect on student achievement (Schneider & Preckel, 2017) and might be especially important in classrooms with a diverse student population (Crose, 2011).

Classroom management is needed for dealing with apparent incivilities, such as talking in class and dominating a discussion (Burke et al., 2014; Nilson, 2010), close supervision and monitoring of students' performances, and using classroom time efficiently (Biggs & Tang, 2011; Kim et al., 2014). Other scholars have debated the relevance of classroom management in higher education on the grounds that students at this level have been in formal education for longer than secondary school students and the student population is more homogenous (William T. Grant Foundation, Spencer Foundation & Bill & Melinda Gates Foundation, 2014). However, in Feldman's (2007) meta-analysis, classroom management is mentioned as a dimension of moderate importance for student achievement ($r = 0.26$).

Many studies have included clear and structured instructions as an essential criterion of quality teaching in higher education (e.g. Feldman, 2007; Hativa et al., 2001). Schneider and Preckel (2017) identified that clear and understandable instructions had the strongest association with student achievement in their category 'presentation' ($d = 1.35$).

There is compelling evidence that confirms the impact of activating teaching on student outcomes. The meta-analysis of Freeman et al. (2014) found that active learning increased students' grades by almost half a standard deviation (0.47) in STEM undergraduate education. Schneider and Preckel's (2017) review study of meta-analyses also highlighted the importance of engaging and interactive teaching, both in lectures and small group settings.

Learning strategies are broadly defined as 'cognitions or behaviours that influence the encoding process and facilitate acquisition and retrieval of new knowledge' (Fryer & Vermunt, 2018, p. 22). As noted by Kirschner et al. (2006, p. 77), 'the goal of instruction . .

. is to give learners specific guidance about how to cognitively manipulate information in ways that are consistent with a learning goal.' This applies particularly to the context of higher education, which is expected to generate strategic, self-regulated, lifelong learners who are able to analyse new information, apply their knowledge to real-world problems, and reflect on proposed strategies (Biggs & Tang, 2011; Ramsden, 2003). Vermetten et al. (1999) provided a longitudinal perspective on the development of learning strategies within higher education. Their results suggest that guidance on the development of student learning strategies impacts students' application of concrete deep-level learning strategies.

Insights derived from studies conducted at lower educational levels have prompted some scholars to argue in favour of a more differentiated instructional approach within higher education to address an increasingly diverse student population (Chamberlin & Powers, 2010; Ernst & Ernst, 2005; Santangelo & Tomlinson, 2009). The few studies examining the impact of differentiated instruction in higher education report promising results, such as higher student satisfaction (Ernst & Ernst, 2005) and a positive impact on student learning (Chamberlin & Powers, 2010; Santangelo & Tomlinson, 2009).

These domains can be classified as foundational or advanced. As described above, without a safe and stimulating classroom climate, efficient organisation of the lesson, and clear instructions, it will be difficult to implement successful activating teaching. Furthermore, studies that focused on the teaching behaviour of student teachers (Van de Grift et al., 2014) and teachers in secondary education (Van der Lans et al., 2018) observed, on average, behaviour in the first three domains (foundational skills), while the latter domains were found to be more challenging (advanced skills). These findings are in line with teaching development theories that describe the advancement of teaching as starting with a focus on oneself as a teacher, focusing on the task, and finally focusing on student outcomes (Akerlind, 2003; Fuller, 1969).

Aim and hypotheses

The present study aims to contribute to the literature by observing to which degree lecturers demonstrate effective teaching behaviour. Following earlier studies in secondary education (Van de Grift et al., 2014; Van der Lans et al., 2018), we hypothesize that lecturers demonstrate more foundational than advanced teaching behaviour. Our second hypothesis is that the number of students in a class impacts teaching behaviour; lecturers teaching small classes demonstrate more advanced teaching behaviour than lecturers teaching large classes. Large classes bring unique challenges to the teaching and learning environment, such as classroom management (Biggs & Tang, 2011; Mueller & Schroeder, 2018), developing students' academic skills (Clarence et al., 2014), and implementing differentiated instruction (Turner et al., 2017).

Methods

Sample

The sample comprises 203 observations. A total of 211 university lecturers were observed once. Of these observations, six were discounted as they entailed student presentations

with little lecturer involvement. One observation was discounted because the observer withdrew from the study, and one observation was deleted because the observation took place in an atypical room (a room in the local cinema). The male–female ratio in the sample is 119:84. Most lecturers (73%) taught a course in the soft sciences compared to 27% in the hard sciences. More bachelor (75%) than master courses (25%) were observed. Courses were typically scheduled for seven weeks. Teaching experience ranged from lecturers who just started teaching to lecturers who had been teaching for 44 years (M = 11.1, SD = 9.1). The class sizes varied between 5 and 220 students (M = 44.9, SD = 45.4).

Observers and observation procedure

Following ethics approval, we sent out invitation letters to all lecturers (over 400) who we knew would be teaching in the second semester (February to June 2017). We did not distinguish between academic rank or type of contract (full-time versus sessional staff). In most faculties, lecturers were asked to voluntarily subscribe to the sample pool. In one faculty, they were informed of the observations and given the option to opt out of the pool.

A total of 25 observers were recruited for the observations. Fourteen educational advisors who mainly focus on professional learning of lecturers were asked to participate as observers. In addition, 11 master students studying in the field of Social Sciences were recruited as observers. Two days of training were organized (one for the educational advisors and one for the students) to increase consensus among the observers and to ensure inter-rater reliability. The training comprised three parts: (a) an explanation and discussion on how to use the instrument and on how to carry out the three-staged observation process consisting of a pre-observation conversation, the actual observation, and a post-observation conversation, (b) observation and debate centring, respectively, on two recorded lessons from a course in the soft sciences and one course in the hard sciences, and (c) logistical information and ethical guidelines.

To match observers with lecturers we created a register in which lecturers' names were replaced by unique numbers. It contained the locations and dates of the observed teaching moments. After observers selected their observations based on the information in the register, they received the lecturer's contact details. Each observer was asked to perform ten observations and to decline those of lecturers whom they knew personally.

Measure

Observation instrument

Authors across secondary and higher education cluster teaching behaviour in similar domains. This similarity in clustering is noticeable in the various observation domains (see Appendix A). In our search for an observation instrument we excluded domains that refer to non-observable teaching behaviour, such as planning and preparation in Danielson's (2013) Framework for Teaching (FTT), and technology in the Teaching Dimensions Observation Protocol (TDOP; Hora et al., 2013).

We opted to use the International Comparative Analysis of Learning and Teaching (ICALT) observation instrument for this study. The ICALT clusters observable teaching

behaviour in six teaching domains and includes a three-item student engagement scale, which is used as a criterion variable. Observers rate all items on a four-point scale, ranging from 1 (mostly weak) to 4 (mostly strong). Although the ICALT has been primarily used in secondary education, we believe the instrument to be suitable for observations in higher education. First, the instrument covers all the concepts applied in the other instruments and comprises items that are based on educational effectiveness research (Hattie, 2008). Second, the instrument reflects a cumulative ordering of teaching skills, corresponding to Fuller's (1969) stage theory of teacher development (Van der Lans et al., 2018). This stage-wise development has also been found in lecturers' teaching conceptions (McLean & Bullard, 2000). Third, the ICALT instrument has been found to be reliable in various settings (Maulana et al., 2016; Van de Grift et al., 2017), demonstrating significant predictive validity between teachers' demonstrated behaviour and students' engagement.

Following Van de Grift et al. (2014), who tailored the original ICALT instrument to the secondary education context (32 items), we adapted this instrument for the higher education context (see Appendix B). First, we tailored its language and formulation. For example, 'The lesson builds up logically going from the simple to the complex' was rephrased as 'The lesson builds up logically (e.g., going from the simple to the complex, comparing different cases etc.)'. Second, we formulated three additional items on the distinctiveness of teaching in higher education based on our literature review. The following two items were added to the domain clear and structured instructions: 'The lecturer presents societal or research developments of the topic' (item 16), and 'The lecturer compares different theories, perspectives or methods' (item 17). Both items were inspired by Marsh's (1982) concept of 'breadth of coverage', referring to how scholarship is conveyed in teaching. The link between teaching and research is a fundamental characteristic of research-oriented higher education institutions. Lecturers should be aware of current developments in their fields 'to support students to realise not only the contested uncertain nature of knowledge, but also the importance and the fascination of pursuing that knowledge' (Healey & Jenkins, 2007, p. 53). The third item, 'The lecturer teaches learners how to reflect' (item 35), was added to the domain teaching learning strategies. Teaching students how to reflect is a universal aim of universities. However, most learners do not spontaneously develop reflection skills. To foster the development of these skills, pedagogical interventions are required across programmes and courses (Ryan & Ryan, 2013, 2013).

Validity

We examined the construct validity of the observation instrument by calculating correlations of the six domains with total scores for teaching skills (Table 1). A lecturer's total score for teaching skills comprised the sum of this individual's scores for the 35 items. The correlations ranged from .42 (differentiation) to .89 (intensive and activating teaching).

The predictive validity of the instrument was considered by correlating the total score for teaching skills with the student engagement scale. There was a strong, positive correlation between the two variables ($r = .61$, $p < .001$). Furthermore, the student engagement scale correlated positively and significantly with all six domains, with values ranging from .32 (safe and stimulating learning climate) to .56 (intensive and activating

Table 1. Descriptive statistics and correlations (Spearman's rho) between domains of teaching behaviour, student engagement scale, and class size (N = 203).

	1	2	3	4	5	6	7	8	Cronbach's α	M	SD	Mean inter-scale correlation
1. Safe and stimulating learning climate	-								.78	3.56	.41	.31
2. Efficient organisation	.49**	-							.45	3.18	.45	.40
3. Clear and structured instructions	.47**	.53**	-						.68	2.83	.49	.48
4. Intensive and activating teaching	.33**	.50**	.68**	-					.77	2.44	.63	.51
5. Differentiation	.06	.15*	.22**	.38**	-				.73	1.32	.47	.23
6. Teaching learning strategies	.20**	.33**	.49**	.64**	.33**	-			.83	2.27	.70	.40
7. Total score for teaching skills	.49**	.61**	.83**	.89**	.42**	.79**	-	.61**	-	90.67	14.44	-
8. Student engagement	.32**	.39**	.47**	.56**	.27**	.50**	-	-	.85	2.92	.66	-
9. Class size[a]	.00	−.23**	−.18**	−.33**	−.08	−.20**	−.25**	−.27**	-	-	-	-

*p <.05 (two-tailed), **p <.01 (two-tailed).
[a]Class size was coded as follows: 0 = small class (0–30 students), 1 = large class (>30 students).

teaching). After applying a Bonferroni correction, the correlations remained significant (p-values of all seven pairs < 0.001).

Reliability

We calculated Cronbach's Alpha for each domain (Table 1). The reliability ranged from .45 (efficient organisation) to .83 (teaching learning strategies), indicating internal consistency for four out of the six domains. Removing item 6 ('the teacher monitors to ensure learners carry out activities in the appropriate manner') increased the Cronbach's Alpha value for the efficient organisation scale from .45 to .63. Deleting item HE1 ('the teachers presents societal or research developments of the topic') increased the Cronbach's Alpha value for the clear and structured instructions scale from .68 to .73. For this descriptive study, we decided to retain items 6 and HE1 with the proviso both here and in the discussion that the results for the respective domains should be interpreted with caution. The reliability of both the total score for teaching skills and the student engagement scale was found to be good (Cronbach's Alpha = .90 and .85, respectively). The mean inter-scale correlation ranged between .23 (differentiation) and .51 (intensive and activating teaching) indicating there is some overlap among the six domains, but they also measure separate aspects of teaching behaviour (Maulana et al., 2016).

Data analysis strategy

After exploring the reliability and validity of the adapted ICALT instrument, we conducted descriptive analyses to examine to which degree lecturers demonstrate effective teaching behaviour. We calculated the mean scores of the teaching domains and used the following criteria: a mean score for a domain that was below 2.5 was considered insufficient, a mean score of 2.5 up to 3.5 was considered sufficient, and a mean score of 3.5 or higher was considered excellent. This classification corresponds to a four-point (1–4) scoring scale for items included in the instrument. When a behaviour was not demonstrated, hardly demonstrated, demonstrated, or very strongly demonstrated, the scoring was 1, 2, 3, and 4, respectively.

To consider the influence of class size on demonstrated teaching behaviour, an independent sample t-test was used to examine how the demonstrated teaching behaviour differs among lecturers who taught classes of varying size. The non-parametric Mann–Whitney U test was used to study how class size impacts the teaching domains that violated the assumption of normality. To categorise classes in small and large classes, we followed the categorisation used by Biggs and Tang (2011). A small class consists of no more than 30 students (n = 116). Large classes hold more than 30 students (n = 87). The class size categories were collapsed for the initial descriptive analyses. Finally, we examined whether lecturers demonstrated a different degree of insufficient, sufficient, and excellent behaviour in small and large classes following the descriptive analyses as explained earlier. All analyses were performed in SPSS 25.0.

Results

Table 2 shows the distribution of the mean score per teaching domain. Most of the lecturers in the sample (68.4%) obtained an excellent mean score in the domain safe and stimulating

Table 2. Observed domain-wise teaching behaviour categorised as insufficient, sufficient, or excellent.

		Criteria (% of lecturers) (N = 203)		
	Domain	Insufficient (< 2.5)	Sufficient (≥2.5–<3.5)	Excellent (≥3.5)
Foundational	Safe and stimulating learning climate	0	31.5	68.4
	Efficient organisation	1.5	66.5	32
	Clear and structured instructions	26.1	65	8.9
Advanced	Intensive and activating teaching	52.7	42.4	4.9
	Teaching learning strategies	58.1	38.5	3.4
	Differentiation	95.1	4.9	0
	Total score for teaching skills	39.4	58.1	2.5

learning climate. None of the lecturers scored insufficient and 31.5% obtained a sufficient mean score. The majority of the lectures scored sufficient in the domains efficient organisation (66.5%) and clear and structured instructions (65%). Whereas almost none of the lecturers (1.5%) obtained an insufficient score in the domain efficient organisation, 26.1% of the lecturers scored insufficient in the domain clear and structured instructions. The reverse is true for the excellent scores in both domains: 32% of the scores in the domain efficient organisation and 8.9% of the scores in the domain clear and structured instructions. Overall, the majority of lecturers demonstrated sufficient behaviour for the instrument's first three domains. These domains reflect foundational teaching skills (Van de Grift et al., 2017).

Observed behaviour relating to the instrument's last three domains (reflecting advanced teaching skills) were comparatively less observed. The scores of 52.7%, 42.4%, and 4.9% of lecturers were, respectively, insufficient, sufficient and excellent for the intensive and activating teaching domain. The results for the teaching learning strategies domain were similar: the respective scores of 58.1%, 38.5% and 3.4% of lecturers were insufficient, sufficient and excellent. Of note are the scores obtained for the differentiation domain. Compared with their scores for other domains, the majority of lecturers (95.1%) attained insufficient scores for this domain and only 4.9% sufficient ones. Finally, although the majority of the observed lecturers obtained a sufficient average total score for teaching skills (58.1%), a relatively large percentage scored insufficiently (39.4%).

To examine whether the differences between observation scores of lecturers teaching small and large classes are significant we performed an independent sample t-test (Table 3). Class size had a significant but small effect on lecturers' observation scores in the domains clear and

Table 3. Results of the independent-sample t-test for four teaching domains and the student engagement domain by class size.

	Class size					
	Small classes (n = 116)	Large classes (n = 87)	95% CI for Mean Difference			
	M (SD)	M (SD)		*t*	*df*	*p*-value
Efficient organisation	3.27 (.46)	3.06 (.41)	.08,.33	3.34	201	.001
Clear and structured instructions	2.90 (.48)	2.73 (.49)	.04,.31	2.51	201	.013
Intensive and activating teaching	2.61 (.60)	2.20 (.59)	.24,.58	4.86	201	.000
Teaching learning strategies	2.39 (.71)	2.11 (.65)	.09,.47	2.89	201	.004
Student engagement	3.06 (.69)	2.72 (.59)	.15,.51	3.65	201	.000

structured instructions (d_s = 0.35), teaching learning strategies (d_s = 0.41), and efficient organisation (d_s = 0.48). The impact of class size was larger on the scores in the domain intensive and activating teaching (d_s = 0.69) as well as on student engagement (d_s = 0.53). In all four domains lecturers teaching small classes scored higher than those teaching large classes. Students in small classes were also more engaged.

A Mann–Whitney *U* test was performed for the two remaining domains, safe and stimulating learning climate and differentiation. Non-significant differences between lecturers were found for both domains (U_{climate} = 5045.5, z = −0.00, p = 1, r = 0.00; $U_{\text{differentiation}}$ = 4615.5, z = −1.14, p = .26, r = 0.08). Both findings were anticipated because of the lack of spread of the scores within these domains.

Conclusion and discussion

In this study we provided a detailed snapshot of the quality of observed teaching behaviour in one semester at a research-oriented university. Our findings indicate lecturers demonstrated teaching behaviour that was deemed sufficient in the three domains reflecting foundational teaching skills. More advanced teaching behaviour, such as intensive and activating teaching and the teaching of learning strategies, was observed less frequently. While lecturers might have the required knowledge related to these domains, they often struggle with implementing this knowledge into practice (Sadler, 2012). This means targeted support is needed to translate knowledge into behaviour.

The two domains that stand out in particular are activating teaching and differentiation behaviour. It is striking that almost half of the observed lecturers did not demonstrate sufficient activating teaching behaviour since the importance of active learning for student achievement has been well-established in higher education (e.g., Freeman et al., 2014). When considering the implications of class size on this behaviour, our results revealed that more activating and intensive teaching behaviour was noticed in small classes. The number of students might make lecturers uncomfortable in trying out active learning methods and the learning space of large classrooms might interfere with organising, for example, group work. That being said, 42.2% of the lecturers teaching small classes demonstrated insufficient activating teaching behaviour. These results are in line with Stains et al. (2018). The second result that stands out is that almost no differentiation behaviour was detected. The absence of this behaviour could be attributed to the fact that lecturers might not be familiar with this concept. Differentiation has only recently been brought to the attention of higher education institutions. Another explanation might be that a course is typically taught over a seven-week period which allows lecturers little time to get to know their students. Had the courses been longer we might have observed more teaching behaviour in the domain differentiation. It is also possible that lecturers differentiated in their teaching but that it was not noticed by the observers because they were unable to identify weaker students or elucidate how the lecturer accounted for inter-learner differences. Furthermore, differentiation also occurs beyond the classroom at the university under study; it is incorporated into course designs or at the programme level, for example, in the honours programme. The honours programme allows top students to follow an additional multidisciplinary programme next to their regular studies.

Documenting teaching behaviour with an observation instrument could support the observer and the observed lecturer by providing them with a common language regarding teaching behaviour. When providing feedback on teaching behaviour, educational

developers should follow the advice of observational frameworks (e.g., Bell, 2001; Bell & Cooper, 2013) and start with a pre-observation conversation in which the goal and the context of the lesson is discussed. Although all teaching behaviour domains impact student learning, the characteristics of the lesson might be a reason for a lecturer to refrain from showing a certain behaviour. When the observation instrument is structurally implemented within an institution, the results can contribute to evaluating the progress of educational innovation projects.

Limitations

As teaching practices vary across higher education institutions, the generalisability of this study's results might be limited due to its data collection at one university in the Netherlands. The scope of the study was further limited in terms of accessing course information and lecturers' intentions. More work needs to be done to develop a full picture of lecturers' teaching behaviour. We will discuss three ideas for future research. First, since preparation and planning are important aspects of teaching quality (Feldman, 2007; Schneider & Preckel, 2017), further studies could explore whether a lecturer had set out to teach in an activating way but had failed to do so. Second, additional data on the place of the observed course in the curriculum, and on whether or not attendance is compulsory, is needed to fully understand the relationship between student engagement and teaching behaviour as observers often indicated that the number of students attending classes was lower than expected. Third, although the observers in this study were trained to ensure interrater reliability, it would be interesting to examine the interrater reliability of the observations if they were to observe in pairs.

Unlike the ICALT instrument used for secondary education, the adapted instrument used in this study has not yet been psychometrically validated. Furthermore, Cronbach's Alpha value ($\alpha = .45$) was low for the efficient organisation domain. At the final meeting with the observers after all observations were completed, the suggestion to further specify examples of good teaching behaviour in the instrument was proposed. For example, 'the lecturer explains clearly which materials can be used' (example of good teaching behaviour for item 7 'provides effective classroom management') could be improved by adding 'the lecturer connects the topic with the self-study materials'. In addition, even though the instrument was adapted for this study, the higher education context could have influenced how observers attributed meanings to the items. For example, 'the lecturer monitors to ensure learners carry out activities in the appropriate manner' (item 6) might have been differently interpreted in the higher education context because of the strong emphasis on students' responsibility for learning and the corresponding perception of the role of a university lecturer. As previously noted, deletion of item 6 increased Cronbach's Alpha value for the scale from .45 to .63. We decided to retain this item because it became apparent during the final meeting with the observers after the observations that it could be especially important for student engagement in large classes during interactive instructions. Nevertheless, the literature suggests that the efficient organisation domain is moderately important for higher education. Therefore, further scrutiny to determine the fit between this domain and the higher education context

is required. In a follow-up study we will draw on the work of Van de Grift et al. (2014) to analyse the psychometric characteristics of the adapted ICALT instrument and to examine the specific stage-wise ordering of teaching behaviour in higher education.

Acknowledgements

We thank all of the observed lecturers and observers for their participation in the study.

Disclosure statement

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

Notes on contributors

Ine Noben is an educational developer and PhD student at the University of Groningen in the Netherlands. Her research interest is in the quality of teaching and professional development in higher education.

Jan Folkert Deinum is a senior educational developer and researcher at the University of Groningen in the Netherlands. His research interest is in the professional development of teachers, academic leadership, and quality assurance in education.

W. H. Adriaan Hofman is a professor of education studies at the University of Groningen in the Netherlands. His research focuses on school and teacher effectiveness, higher education, research methods, and education in developing countries.

ORCID

Ine Noben http://orcid.org/0000-0003-2019-9224

References

Akerlind, G. S. (2003). Growing and developing as a university teacher: Variation in meaning. *Studies in Higher Education*, *28*(4), 375–390. https://doi.org/10.1080/0307507032000122242

Alicea, S., Suárez-Orozco, C., Singh, S., Darbes, T., & Abrica, E. J. (2016). Observing classroom engagement in community college: A systematic approach. *Educational Evaluation and Policy Analysis*, *38*(4), 757–782. https://doi.org/10.3102/0162373716675726

Bell, M. (2001). Supported reflective practice: A programme of peer observation and feedback for academic teaching development. *International Journal for Academic Development*, *6*(1), 29–39. https://doi.org/10.1080/13601440110033643

Bell, M., & Cooper, P. (2013). Peer observation of teaching in university departments: A framework for implementation. *International Journal for Academic Development*, *18*(1), 60–73. https://doi.org/10.1080/1360144X.2011.633753

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Open University Press.

Burke, L. A., Karl, K., Peluchette, J., & Randy Evans, R. (2014). Student incivility: A domain review. *Journal of Management Education*, *38*(2), 160–191. https://doi.org/10.1177/1052562913488112

Chamberlin, M., & Powers, R. (2010). The promise of differentiated instruction for enhancing the mathematical understandings of college students. *Teaching Mathematics and Its Applications*, *29*(3), 113–139. https://doi.org/10.1093/teamat/hrq006

Clarence, S., Albertus, L., & Mwambene, L. (2014). Building an evolving method and materials for teaching legal writing in large classes. *Higher Education*, *67*(6), 839–851. https://doi.org/10.1007/s10734-013-9707-8

Crose, B. (2011). Internationalization of the higher education classroom: Strategies to facilitate intercultural learning and academic success. *International Journal of Teaching and Learning in Higher Education*, *23*(3), 388–395. https://www.isetl.org/ijtlhe/pdf/IJTLHE1056.pdf

Danielson, C. (2013). *The framework for teaching: Evaluation instrument*. The Danielson Group.

Ernst, H. R., & Ernst, T. L. (2005). The promise and pitfalls of differentiated instruction for undergraduate political science courses: Student and instructor impressions of an unconventional teaching strategy. *Journal of Political Science Education*, *1*(1), 39–59. https://doi.org/10.1080/15512160590907513

Feldman, K. A. (2007). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 93–144). Springer.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *PNAS*, *111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Fryer, L. K., & Vermunt, J. D. (2018). Regulating approaches to learning: Testing learning strategy convergences across a year at university. *British Journal of Educational Psychology*, *88*(1), 21–41. https://doi.org/10.1111/bjep.12169

Fuller, F. F. (1969). Concerns of teachers: A developmental conceptualization. *American Educational Research Journal*, *6*(2), 207–226. https://doi.org/10.3102/00028312006002207

Hativa, N., Barak, R., & Simhi, E. (2001). Exemplary university teachers: Knowledge and beliefs regarding effective teaching dimensions and strategies. *The Journal of Higher Education*, *72*(6), 699–729. https://doi.org/10.1080/00221546.2001.11777122

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hattie, J. (2015). The applicability of visible learning to higher education. *Scholarship of Teaching and Learning in Psychology*, *1*(1), 79–91. https://doi.org/10.1037/stl0000021

Healey, M., & Jenkins, A. (2007). Strengthening the teaching-research linkage in undergraduate courses and programs. *New Directions for Teaching and Learning*, *2006*(107), 43–53. https://doi.org/10.1002/tl.244

Holley, L. C., & Steiner, S. (2005). Safe space: Student perspectives on classroom environment. *Journal of Social Work Education*, *41*(1), 49–64. https://doi.org/10.5175/JSWE.2005.200300343

Hora, M. T., & Ferrare, J. J. (2013). Instructional systems of practice: A multidimensional analysis of math and science undergraduate course planning and classroom teaching. *Journal of the Learning Sciences*, *22*(2), 212–257. https://doi.org/10.1080/10508406.2012.729767

Hora, M. T., Oleson, A., & Ferrare, J. J. (2013). *Teaching dimensions observation protocol (TDOP): User's manual*. Wisconsin Center for Education Research.

Jacob, J. W., Xiong, W., & Ye, H. (2015). Professional development programmes at world-class universities. *Palgrave Communications*, *1*(15002), 1–27. https://doi.org/10.1057/palcomms.2015.2

Kim, M. K., Kim, S. M., Khera, O., & Getman, J. (2014). The experience of three flipped classrooms in an urban university: An exploration of design principles. *The Internet and Higher Education*, *22*, 37–50. https://doi.org/10.1016/j.iheduc.2014.04.003

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Marsh, H. W. (1982). SEEQ: A reliable, valid, and useful Instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology*, *52*(1), 77–95. https://doi.org/10.1111/j.2044-8279.1982.tb02505.x

Maulana, R., Helms-Lorenz, M., & Van de Grift, M. (2016). Validating a model of effective teaching behaviour of pre-service teachers. *Teachers and Teaching*, *23*(4), 471–493. https://doi.org/10.1080/13540602.2016.1211102

Maulana, R., Helms-Lorenz, M., & Van de Grift, W. (2015). Development and evaluation of a questionnaire measuring pre-service teachers' teaching behaviour: A Rasch modelling approach. *School Effectiveness and School Improvement*, *26*(2), 169–194. https://doi.org/10.1080/09243453.2014.939198

McLean, M., & Bullard, J. E. (2000). Becoming a university teacher: Evidence from teaching portfolios (how academics learn to teach). *Teacher Development*, *4*(1), 79–101. https://doi.org/10.1080/13664530000200104

Mueller, R., & Schroeder, M. (2018). From seeing to doing: Examining the impact of non-evaluative classroom observation on teaching development. *Innovative Higher Education*, *43*(5), 397–410. https://doi.org/10.1007/s10755-018-9436-0

Nilson, L. B. (2010). *Teaching at its best. A research-based resource for college instructors* (3th ed.). Jossey-Bass.

O'Malley, K. J., Moran, B. J., Haidet, P., Seidel, C. L., Schneider, V., Morgan, R. O., Kelly, P. A., & Richards, B. (2003). Validation of an observation instrument for measuring student engagement in health professions settings. *Evaluation & the Health Professions*, *26*(1), 86–103. https://doi.org/10.1177/0163278702250093

Pianta, R. C., Hamre, B. K., & Mintz, S. L. (2012). *The CLASS-secondary manual*. University of Virginia.

Postareff, L., Lindblom-Ylänne, S., & Nevgi, A. (2008). A follow-up study of the effect of pedagogical training on teaching in higher education. *Higher Education*, *56*, 29–43. https://doi.org/10.1007/s10734-007-9087–z

Ramsden, P. (2003). *Learning to teaching in higher education* (2nd ed.). RoutledgeFalmer.

Ryan, M. (2013). The pedagogical balancing act: Teaching reflection in higher education. *Teaching in Higher Education*, *18*(2), 144–155. https://doi.org/10.1080/13562517.2012.694104

Ryan, Ma., & Ryan, Mi. (2013). Theorising a model for teaching and assessing reflective learning in higher education. *Higher Education Research & Development*, *32*(2), 244–257. https://doi.org/10.1080/07294360.2012.661704

Sadler, R. (2012). The challenges for new academics to adopting student-centred approaches to teaching. *Studies in Higher Education*, *37*(6), 731–745. https://doi.org/10.1080/03075079.2010.543968

Santangelo, T., & Tomlinson, C. A. (2009). The application of differentiated instruction in postsecondary environments: Benefits, challenges, and future directions. *International Journal of Teaching and Learning in Higher Education*, *20*(3), 307–323. http://www.isetl.org/ijtlhe/pdf/IJTLHE366.pdf

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Sciences and Mathematics*, *102*(6), 245–253. https://doi.org/10.1111/j.1949-8594.2002.tb17883.x

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin*, *143*(6), 565–600. https://doi.org/10.1037/bul0000098

Smith, M. K., Jones, F. H. M., Gilbert, S. L., & Wieman, C. E. (2013). The classroom observation protocol for undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE—Life Sciences Education*, *12*(4), 618–627. https://doi.org/10.1187/cbe.13-08-0154

Stains, M., Barker, J., Chasteen, H. M. K., Cole, S. V., DeChenne-Peters, R., Eagan, S. E., Jr., Esson, M. K., Knight, J. M., Laski, J. K., Levis-Fitzgerald, F. A., Lee, M., Lo, C. J., McDonnell, S. M., McKay, L. M., Michelotti, T. A., Musgrove, N., Palmer, A., Plank, M. S., Rodela, K. M., & Young, M. H. (2018). Anatomy of STEM teaching in North American universities. *Science Education*, *259*(6383), 1468–1470. https://doi.org/10.1126/science.aap8892

Stes, A., Min-Leliveld, M., Gijbels, D., & Van Petegem, P. (2010). The impact of instructional development in higher education: The state-of-the-art of the research. *Educational Research Review*, *5*(1), 25–49. https://doi.org/10.1016/j.edurev.2009.07.001

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. *Teaching in Higher Education*, *22*(7), 822–838. https://doi.org/10.1080/13562517.2017.1301907

Trigwell, K., & Prosser, M. (2004). Development and use of the approaches to teaching inventory. *Educational Psychology Review*, *16*(4), 409–424. https://doi.org/10.1007/s10648-004-0007-9

Turner, W. D., Solis, O. J., & Kincade, D. H. (2017). Differentiating instruction for large classes in higher education. *International Journal of Teaching and Learning in Higher Education*, *29*(3), 490–500. https://www.isetl.org/ijtlhe/pdf/IJTLHE2735.pdf

Van de Grift, W. (2007). Quality of teaching in four European countries: A review of the literature and application of an assessment instrument. *Educational Research*, *49*(2), 127–152. https://doi.org/10.1080/00131880701369651

Van de Grift, W., Chun, S., Maulana, R., Lee, O., & Helms-Lorenz, M. (2017). Measuring teaching quality and student engagement in South Korea and The Netherlands. *School Effectiveness and School Improvement*, *28*(3), 337–349. https://doi.org/10.1080/09243453.2016.1263215

Van de Grift, W., Helms-Lorenz, M., & Maulana, R. (2014). Teaching skills of student teachers: Calibration of an evaluation instrument and its value in predicting student academic engagement. *Studies in Educational Evaluation*, *43*, 150–159. https://doi.org/10.1016/j.stueduc.2014.09.003

Van der Lans, R. M., Van de Grift, W., & Van Veen, K. (2018). Developing an instrument for teacher feedback: Using the Rasch model to explore teachers' development of effective teaching strategies and behaviors. *Journal of Experimental Education*, *86*(2), 247–264. https://doi.org/10.1080/00220973.2016.1268086

Vermetten, Y. J., Vermunt, J. D., & Lodewijks, H. G. (1999). A longitudinal perspective on learning strategies in higher education – Different view-points towards development. *British Journal of Educational Psychology*, *69*(2), 221–242. https://doi.org/10.1348/000709999157699

William T. Grant Foundation, Spencer Foundation, & Bill & Melinda Gates Foundation (2014). *Measuring instruction in higher education: Summary of a Convening*. http://wtgrantfoundation.org/new-report-measuring-instruction-in-higher-education