Total Quality Management, 2016
Vol. 27, No. 9, 1063–1090, http://dx.doi.org/10.1080/14783363.2015.1060850

ClassroomQual: a scale for measuring the use-of-classrooms-for-teaching–learning service quality

Jorge Pérez Rave* and Leandro Muñoz Giraldo

Department of Industrial Engineering, University of Antioquia, Medellin, Colombia

This study has two purposes: (1) to develop and refine a measure of the quality of the complementary service named 'use-of-classrooms-for-teaching–learning' and (2) to contribute to the empirical evidence regarding the dimensions of the complementary service. The design was cross-sectional of development and scale validation. The research methods were based on factor and psychometric analyses. The study was conducted in three phases. In the initial phase, 30 interviews were conducted in order to collect critical incidents. The second phase involved the improvement of items (self-administered questionnaire of 59 items, Cronbach's α, and exploratory factor analyses). Students from Industrial Engineering at the University of Antioquia comprised the sample ($n = 290$). In the third phase, the confirmatory factor analysis (CFA) was implemented, using Amos 21.0 software. The items that made it through the exploratory stage was 23, classified under 7 latent factors (the initial model in the CFA). Through strategies of re-specification, four more models were arrived at, and finally two models (a 16-item model and a 15-item one), both with seven latent factors, were chosen. These models are two versions of a scale we have called ClassroomQual, which presents satisfactory measures of overall fit, parsimony, reliability, and validity.

Keywords: confirmatory factor analysis; complementary services; service quality; ClassroomQual; higher education

1. Introduction

Several studies have found a significant positive relationship between student perceived quality, satisfaction, and retention (Astin, 1993; Temtime & Mmereki, 2011; Zhao, 2012). This finding makes evaluation and enhancement of student perception of service quality a key factor impacting on the level of success of higher education institutions. However, from the student's perspective, this subject is still in its infancy, despite the fact that there are several studies on it (Shank, Walker, & Hayes, 1995; Zhao, 2012). So, to date, there is no measurement instrument available that captures the complexities, as well as all the factors related to the service, from the student's perspective, and there are several reasons for this.

First, there are unknown dimensions that comprise service quality from the student perspective (Abdullah, 2006a, 2006b; Annamdevula & Bellamkonda, 2012; Dlačića, Arslanagićb, Kadić-Maglajlićb, Markovićc, & Raspord, 2014; Mohamad, Hassan, Rahman, & Ghouri, 2012; Zhao, 2012), which might be explained by the trend towards design of quality assurance indicators into the processes, from the provider perspective (Kadhim, Taqi, & Shuaibu, 2012). Second, most of the dimensions that are currently measured in higher education have been derived from business (such as the SERVQUAL dimensions), and are not necessarily applicable (Barani & Kumar, 2013; Chua, 2004;

*Corresponding author. Email: jorge.perez@udea.edu.co

Koni, Zainal, & Ibrahim, 2013). Third, the few studies that have identified critical incidents (CI) of the service, from the higher education context (i.e. HedPERF), still do not allow for generalisation. Therefore, some authors (Icli & Anil, 2014; Law, 2013) highlight the need to consider the particularities as well as the contextual complexities that have been disregarded in mainstream studies to date. In this regard, there are authors who point out the need to study not only the 'service product' but also, independently, the 'service delivery' (Miller & Brooks, 2010). This last category involves professor, classroom (or virtual environments), teaching–learning methodologies, and communication systems, among others. So, the point is that understanding and enhancing the core service of teaching–learning are not enough (Mohamad et al., 2012) without considering, among others, the complementary services, especially since these aid the performance of the core service, and if they fail, the purpose of the core service fails as well (Grönroos, 1992; Miller & Brooks, 2010; Nóbrega, Mendonca, De Souza, & Campos, 2010; Sheeja, Krishnaraj, & Harindranath, 2014).

Despite this call for understanding students' whole experience regarding higher education service quality (Chen, Huang, Shu, & Wang, 2013; Dlačića et al., 2014; Hittman, 1993; Miller & Brooks, 2010), there is scarce research that has focused on it. Instead, most of the literature continues to focus on the academic aspects (Mohamad et al., 2012). Indeed, in the best of the cases, when these studies encompass the complementary services, they are considered from a generalist perspective, including them only as quality dimensions. This ignores another challenge highlighted in the literature: discovering and deepening particularities of new service typologies, customers, and contexts (Annamdevula & Bellamkonda, 2012; Icli & Anil, 2014; Muñoz & Pérez, 2013).

Considering the above, the present study joins the efforts of the academic community in order to provide a better understanding of higher education service quality, and in this regard, developing the research in the Engineering Faculty of the University of Antioquia (Medellín-Colombia), the objective of the study is to obtain and refine a measurement instrument (scale) to assess the quality, from the student perspective, of the complementary service (not the traditional core service: teaching–learning) that takes place 'during class', which has been named 'use-of-classrooms-for-teaching–learning'.

Among the different complementary services at the University of Antioquia, 'the use-of-classrooms-for-teaching–learning' service has been selected as the object of analysis, for three reasons. First, due to its proximity to the core service: this service provides the facilities for performing the class, the technological means, and the assistance for possible arising eventualities, among others. The support department in the Engineering Faculty, which may be considered the service provider, delivers this complementary service. The second reason is some students' (as well as the service provider's) assertions regarding the possibilities of bettering the complementary service, as it currently affects the core service quality. For instance, if the projector does not work at the beginning of or during class, student concentration is affected negatively. Valuable time is wasted waiting for the provider to arrive in the classroom (the provider sends out support personnel). Indeed, on occasion the upset can be such that students have to move to another classroom. The third reason relates to the appropriateness of deepening and divulging this service, given such context particularities as detachment of fully identified providers from the university academic units they serve. In fact, the reviewed literature failed to show antecedents or a denomination for this service.

The second objective of the study is to contribute to the empirical evidence regarding the inherent subtleties of the complementary service, such as its quality structure and dimensions. This end makes it possible to analyse whether the dimensions are similar to the traditional, or whether they are contextual particularities, upon which some authors (Annamdevula & Bellamkonda, 2012; Icli & Anil, 2014; Muñoz & Pérez, 2013) suggest further research is needed.

Regarding managerial implications, the 'use-of-classrooms-for-teaching–learning' service quality evaluation as well as the measurement instrument constitute a means of support for delivering a better value-offering for the main customers: students.

The remainder of the article is organised as follows: Section 2 presents the literature review; Section 3 details the materials and methods employed; Section 4 introduces the results obtained. Section 5 discusses the significance of the results, in consideration of the context where the instrument was applied. Reference sources are provided. Section 6 presents the general conclusions of the study. Section 7 addresses the study's limitations while advising on future research.

2. Literature review

In the current global environment, service quality is a critical factor for competitiveness (Abdullah, 2006a; Marshall & Murdoch, 2001), and the field of education is no stranger to this trend, precisely, among other aspects, due to the increasing interest in an internationalised education (Mohamad et al., 2012; de Rijke & Plucker, 2011; Yeo & Li, 2012; Yildiz, 2014; Zhao, 2012).

Despite the widespread success of service quality studies in the business sector, their use in education is still in its infancy stage (Shank et al., 1995; Zhao, 2012). On 8 January 2014, for example, 3136 papers were found, where the term 'service quality' was searched on Scopus. However, only 73 papers (2.3%) were recovered, where the following algorithm was executed: Title ('service quality') and Title-Abs-Key ('higher education'). More evidence can be found in a previous systematic review by Pérez and Muñoz (2014), where only 3 of the 47 papers reviewed focused on education services (6.3%). Most studies focused on e-service (20%) and hospitality service (16%), which, in accordance with global trends, belong in ICT and tourism.

Several authors agree that a difficulty in measuring service quality in higher education is that the definition of service quality in higher education remains unclear (Koni et al., 2013; Reeves & Bednar, 1994). However, they recommend considering such assumptions about the nature of overall service quality as customer focus (Kessler, 1995). Another difficulty in the field is the complexity of identification of service customers. This is mainly because of the different service typologies offered. Furthermore, different services are demanded by several kinds of customers. It is worth noting that students are the main protagonists in the current educational system (Quinn, Lemay, Larsen, & Johnson, 2009).

Additionally, higher education institutions have tended to focus on the management of teaching–learning processes while adopting such initiatives of quality assurance as accreditation programmes (Abdullah, 2006a, 2006b). Although such programmes are essential, they are not sufficient to represent the many factors and complexities involved in education service quality. In this regard, Kadhim et al. (2012) argue that to assure higher education service quality, institutions have largely turned to the design of indicators from the provider perspective (e.g. educators), giving little attention to the student perspectives. However, recently, researchers have started to show a growing interest in the latter approach (Abdullah, 2006a, 2006b; Dlačića et al., 2014; Koni et al., 2013; Miller & Brooks, 2010).

A higher education service quality study from the student's viewpoint is essential, among other aspects, for the survival and success of institutions, in the mid and long term, due to general agreement across research studies on the positive relation between perceived quality, satisfaction, and retention (Astin, 1993; Temtime & Mmereki, 2011; Zhao, 2012). It can, therefore, be assumed that if students are satisfied with the service, they are less likely to look to change service providers. Moreover, students tend to advertise the image of their institution, as they perceive it, via word of mouth (Jiewanto, Laurens, & Nelloh, 2012).

In order to improve service quality, education institutions have normally used techniques and strategies that have traditionally been applied in business contexts (Barani & Kumar, 2013; Chua, 2004; Koni et al., 2013). Considering the premise 'that which is not measured can not be improved', measurement of higher education service quality has justly been influenced by the SERVQUAL instrument (Koni et al., 2013; Rasli, Shekarchizadeh, & Iqbal, 2012; Sorayael, Ahangar, Kasiri, & Baharami, 2013; Zhao, 2012; Zulkefli & Uden, 2013). SERVQUAL has been applied to business for over 20 years (Chen et al., 2013; Yildiz, 2014). This instrument (Parasuraman, Berry, & Zeithaml, 1991; Parasuraman, Zeithaml, & Berry, 1988) assesses service quality by measuring gaps between the expectations and perceptions of customers. The initial version consisted of 10 dimensions, but in a subsequent study, the number of dimensions was reduced to 5: tangibles, reliability, responsiveness, assurance, and empathy.

Despite the rising popularity of SERVQUAL, its critics question the relevance of the expectations component, arguing that it lacks theoretical support, which in turn may turn into increased bias. Consequently, Cronin and Taylor (1992) proposed SERVPERF, which eliminated the expectations component of SERVQUAL, focusing solely on the measurement of perceptions. However, to date, there is little agreement as to which instrument performs better (Yildiz, 2012). Each of the above instruments has contributed to service quality assessment in various fields and has inspired the development of similar instruments.

Although service quality measurement has evolved through SERVQUAL, SERVPERF, and other generic instruments, various authors (Abdullah, 2006b; Firdaus, 2006; Quinn et al., 2009) highlight that instruments such as these are not designed to pick up all of the subtleties associated with education service, especially as this service has inherent characteristics that differentiate it from other types of services.

One argument for the above is the fact that SERVQUAL does not consider the academic component in all its depth, as such level of detail is more appropriate to those services which deal more with personal interaction than with technical issues (Chen et al., 2013). Additionally, Zhao (2012) and Barani and Kumar (2013) mention that results obtained through SERVQUAL cannot be generalised to all education contexts, particularly when the service types proposed by Grönroos (1992) (core and complementary services) are being considered. Core service refers to the product that justifies the existence of a company. Complementary services are divided into facilitating and support services. Facilitating as a service ensures that the core service is provided. On the other hand, support service is provided to increase core service value and differentiate it from that of competitors. Nóbrega (1997) elaborates on this position by arguing that the classification of service into 'facilitating' and 'support' may cause confusion. Therefore, he proposes the terms 'complementary' and 'supplementary' services, respectively.

According to Nóbrega et al. (2010), in the higher education context, core service corresponds to the class (teaching–learning dimension), while complementary services can be understood as 'Directors' attendance, library, central customer service, cleaning and building maintenance, snack services, surveillance services and reprographic services.'

On the other hand, the supplementary services are 'self-service option, service portal, information services and computer labs services' (p. 10). Other authors argue that the few studies on service quality measurement from the students' viewpoint have tended to focus on academic aspects, exclusively (Mohamad et al., 2012). However, the study of complementary and supplementary services is required to understand the overall student experience during service (Abdullah, 2006b; Chen et al., 2013; Dlačića et al., 2014; Hittman, 1993).

Responding to the need of focusing solely on the higher education sector, Firdaus (2006) proposed the HedPERF (Higher education PERFormance-only) scale. This instrument is composed of 41 items, which are grouped into 6 dimensions: non-academic aspects, academic aspects, reputation, access, programmed issues, and understanding. Assuming a generalist perspective in the service quality evaluation, the scale HedPERF gathers several aspects of the experience of different service typologies by the customer (within the higher education context). However, this results in a reduced understanding of the complexities inherent to each typology. For instance, regarding complementary services that take place 'during class', the scale only considers one item for evaluation (minimal class sizes).

In a comparison between HedPERF and other generic instruments on service quality measurement, Abdullah (2006a) found, in a specific educational context, that HedPERF was more reliable and valid than SERVPERF. However, these same authors mention that this finding cannot be generalised to all educational contexts. Law (2013) has argued a different position. In one specific context, he had identified that all of SERVPERF dimensions were valid, while only three of HedPERF were.

Following Firdaus (2006), to date, there are several studies about higher education service quality that adopt this generalist perspective. Icli and Anil (2014), for instance, developed the HEDQUAL scale for MBA programmes, from the student perspective. The scale contains 26 items grouped into five quality dimensions: academic quality, administrative services quality, library services quality, support services quality, and careers advisory services quality. However, only 2 out of these 26 items relate to elements of the complementary services that occur 'during class': (1) size of the classrooms and (2) necessary classroom equipment (computer, digital projector, etc.). Therefore, certain particularities that underlie this service are not taken into account (i.e. technological equipment maintenance services, classroom environment conditions, and classroom physical adequacy, among others).

Similarly, Annamdevula and Bellamkonda (2012) proposed the HiEdQUAL scale to measure higher education service quality in an institution in India. The HiEdQUAL is composed of 27 items grouped into 5 quality dimensions: teaching and course content, administrative services, academic facilities, campus infrastructure, and support services. Only 3 out of the 27 items actually evaluate complementary services aspects 'during class': (1) classrooms equipped with teaching aids, (2) sufficient classroom availability, and (3) computer labs with adequate equipment and Internet connectivity.

On the other hand, focusing on the student experience 'during class' (teaching–learning), which can be studied as the service (Schneider, Hanges, Goldstein, & Braverman, 1994; Stodnick & Rogers, 2008), Miller and Brooks (2010) also argue that current service quality instruments do not sufficiently recognise such service. This is critical as much evidence shows that perceptions among professors and students, regarding student expectations, are not necessarily the same (Miller & Brooks, 2010; Shank et al., 1995). This is in line with the repeated challenges by service quality research, which highlight the need to study each higher education service context thoroughly. All these are necessary in order to identify, evaluate, and understand the dimensions and determinant factors for service quality (Abdullah, 2006a, 2006b; Dlačića et al., 2014; Mohamad et al., 2012; Zhao, 2012).

Adopting a focused outlook, with a minor scope regarding the services variety, which differs from the traditional perspective, Miller and Brooks (2010) propose the ClassQual scale. This is a scale that contains 13 items grouped into two quality dimensions: service delivery and service product. ClassQual focuses specifically on the student experience 'during class', in an effort to encompass the complexities that take place at such a moment of the service. However, regarding service delivery, the scale just considers elements related to a professor's competencies such as the ability to communicate, the level of respect shown to the students, and the professional appearance, among others. Therefore, other provider resources (complementary services), such as classroom materials, technological devices, and support service staff, among others, which might take place during service delivery, are not considered.

In addition to the above perspective, it is worth mentioning the instrument named Student Evaluation of Educational Quality (SEEQ), proposed by Marsh (1987). This 33-item instrument also focuses on the student experience regarding the class, but under nine dimensions: learning, enthusiasm, organisation, group interaction, personal rapport, breadth of coverage, examinations, assignments, and overall. This scale is widely used and outstanding in terms of its reliability and validity, among others. As in the case of Miller and Brooks (2010), the scope of SEEQ is bounded to the academic aspects (course, instruction, teaching methods, assessment, etc.) and not to the complementary elements that can take place during the core of service deployment.

In keeping with the above, the present study differs significantly from Miller and Brooks (2010), while still complementing it, in order to examine the wider set of elements that comprise service delivery.

Therefore, despite the progress in higher education service quality research, there are many gaps and discrepancies in it that need to be resolved. To date, there is no consensus on the structure of the service. Neither is there an instrument to measure all the complexities of it. This, in turn, leads to insufficient evidence on the perceptions and expectations of the service by various customer types.

In consequence, this paper aims to join efforts with researchers in the service quality field. In order to achieve this, we have studied a complementary service named 'use-of-classrooms-for-teaching–learning', in order to develop and refine a scale for the measurement of the service, not focusing, as it has been traditionally, on the core service: the teaching–learning process itself. On the contrary, this study turns to the 'use of classroom' experience during service delivery.

3. Materials and methods

3.1. *The use-of-classrooms-for-teaching–learning service*

In the context of this study, students experience the service each time they arrive in the classroom, take it, utilise it, and finally, leave it. In the service sequence, 'classroom' is understood as the set of materials, informatics, and the environment that exists inside the enclosure, for the teaching–learning process (projector, whiteboard, floor, seats, PC, desk, and environment). In the context of the study, the 'classroom' described is set up in the traditional classroom layout for lectures, and does not involve other sorts of specialised facilities such as laboratories (physics, chemistry, etc.) and auditoriums, among others.

The experience of the students regarding the service starts on arriving in the classroom. At this moment, they take it and proceed with its use, in order to facilitate the teaching–learning process. Once the class ends, they leave the classroom and the service finishes. The moment of truth associated with the arrival, use, and departure from the classroom is mediated by a human resource that is responsible for opening and delivering the classroom, in addition to managing the technological needs that arise during the teaching–learning process (i.e. device set-up, device performance, and troubleshooting, among others).

3.2. *Design and research methods*

The research design was one of cross-sectional development and validation of scales. The research methods were based on factor and psychometric analyses. The procedure was carried out in three phases. The first consisted of obtaining the assessment items, using the CI technique, and review of the literature. The second phase involved the refinement of items, using a survey in the form of a structured self-administered questionnaire, as well as exploratory factor analysis (EFA) and Cronbach α reliability analysis. The third phase consisted of the final validation of the scale using confirmatory factor analysis (CFA).

3.3. *Procedure*

3.3.1. *Item generation*

The initial items were obtained through literature review and interviews, with open questions, following the method of CI (Hayes, 2002). Thirty students from different academic levels were interviewed, collecting positive and negative aspects of the use-of-classroom-for-teaching–learning service from the students' viewpoint. The open question administered was 'what do you like best and least about the service', having previously offered an explanation of the moments of truth of the service cycle, by means of the use of visual cues. Every time a vague answer was submitted, other questions that helped specify it were used, as recommended by Hayes (2002). For example, if someone responded, 'I like it because it's good', the person would then be asked, 'what do you specifically mean by "good"?' Or, 'specifically, what makes a service "good"?' In the end, 268 CI were collected. Later, these were grouped into satisfaction elements (SE), which constituted the questionnaire items. The moments of truth of the service and the absolute and relative frequencies of the SE are presented in Figure 1.

It is worth mentioning that most of the elements (87.4%) refer to the third moment of truth, named 'using the classroom'.

3.3.2. *Item refinement*

Firstly, a cross-sectional study with a structured self-administered questionnaire was carried out. The questionnaire consisted of 61 items, 59 of which corresponded to the construct of interest resulting from the previous phase. All the items were measured on a 7-point Likert scale (1 = 'strongly disagree' and 7 = 'strongly agree'). The participants in this study were students of the Industrial Engineering Department at the University of Antioquia in Medellin, Colombia, who voluntarily completed an initial questionnaire just before the beginning of classes. Because of missing data and incomplete data, the final sample size was 290. This was distributed as follows: 43% female and 57% male. Most of the students (84%) were between ages 17 and 23 years old. Also, the sample included students from all academic levels (based on the grade point average [GPA]),

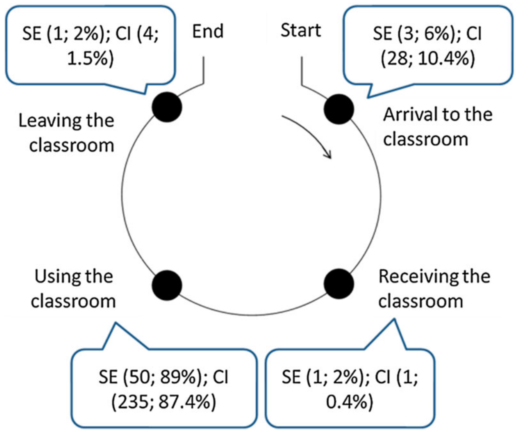

Figure 1. The service cycle.

ranging from 15% to 18% (academic level 1: 3–5 GPA), from 6% to 8% (academic level 2: 6–8 GPA), and from 3% to 4% (academic levels 3: 9–10 GPA). Finally, it should be noted that the sample represented 37.8% of the active population by 15 July 2013 (the total number of students registered in the Industrial Engineering Department was 767), in this way overcoming the 95% confidence level at 5% estimated error for '*p*' (256), as required by proportion-based statistical sampling.

The sample selection was derived via purposive sampling. The actual sample included only students of the Industrial Engineering programme, due to the following reasons: (1) the scope of the work was not to study the possible differences among groups; (2) the industrial engineering programme is one of the largest in the faculty, facilitating the election of a sample large enough for the purpose of the study: design and validation of the measurement instrument; (3) the easiness of access to students, according to established criteria; and (4) the fact that rest of the faculty student population, besides sharing the same classrooms for lectures, require laboratories (chemistry, mechanics, electronics, etc.) administered by other providers. The latter numeral is important in order to obtain a more homogeneous sample and diminish the evaluation bias.

The EFA was run on 59 items, using the SPSS software, to identify the initial structure of the service quality studied. The EFA was run through the principal component extraction, using Varimax rotation. As a refining criterion, only items with factor loadings above 0.45 were accepted (Hair, Anderson, & Tatham, 1998). To obtain the optimal number of factors in the final solution, the latent root criterion was used: only factors with eigenvalues greater than 1 were considered (scree test) (Hair et al., 1998).

After this first filter, 14 factors with eigenvalues greater than 1 were identified, and 10 items presenting loads lower than 0.45 in all factors were eliminated. After this procedure was repeated, 12 factors were carried over and 3 more items were eliminated, 2 of them because they did not have the minimum load and the other because it showed high loads on more than 1 factor. The results of the EFA were complemented by measuring the Cronbach α coefficients (performed in SPSS 14.0), as well as by evaluating the reliability of each construct and their behaviour once items had been removed. Five of the 12 factors did not reach the minimum value required for reliability (minimum 0.7). Three of them were lower than 0.59 and the remaining two were 0.65 and 0.68. It should be noted that, after deleting an item, the change in the value of each factor was analysed by the Cronbach α coefficient, but there was no improvement in reliability in any of the cases. Seven factors were arrived at: one 9-item factor, two 5-item factors, one 3-item factor, and three 2-item factors, for a total of 28 items.

In the three most representative factors, some items were found to be redundant in their semantic content. The decision of which ones to keep was made by their face validity (Thompson, Haddad, & Smith, 2014) supplemented by an item-total correlation analysis (less than 0.4, according to Mattick & Clarke, 1998), and a Cronbach α reliability analysis. This led to 4 items being discarded, bringing it to a total of 24 items. EFA was run on these 24 items, producing a 7-factor structure again, while discarding one more item because of no significant load on any factor. Finally, EFA was run on the remaining 23 items, and the 7-factor structure was maintained, explaining 68.8% of the variance, as shown in Table 1.

Table 2 shows the definition of the factors, the items that make them up and their corresponding factor loads after EFA, and the Cronbach α reliability rates. The reliability values for the identified factors ranged between 0.71 and 0.88, meeting the criterion of the minimum value of 0.70 (Nunnally, 1987).

3.3.3. *Final validation*

The purpose of this phase is to confirm and refine the preliminary scale. In this way, the initial model, derived from the final EFA is evaluated (7 latent factors, 23 items, see Table 2), seeking to ascertain whether it reasonably fits the empirical evidence. In order to achieve this, fit (χ^2 ratio/df, standardized root mean square residual (SRMR), and comparative fit index (CFI), among others), parsimony (Akaike Information Criterion [AIC]), and replicability indices (expected cross-validation index [ECVI]) are noted. These indices may be consulted in several CFA studies, including Rial, Valera, Abalo, and Lévy (2006). If the global model fit indices, its validity, item reliability, and modification indices suggest modifications, re-specification of the initial model is required, leading to one or more alternative models, used for comparison. The chosen model (the top performer by the aforementioned indices) is then analysed in terms of its convergent validity, discriminant validity, and reliability. Finally, the structural model is validated, considering the significance of its parameters.

3.3.4. *Global model evaluation*

Amos 21 (SPSS) was used to conduct the CFA. The application of CFA was carried out using a six-step approach (Lam & Bae, 2013): (1) model specification, (2) identification,

Table 1. Factors with eigenvalues greater than 1 and the total explained variance.

	Initial eigenvalues			Sum of the squared rotation loadings		
Component	Total	% Variance	% Cumulative	Total	% Variance	% Cumulative
1	5.6	24.419	24.419	4.164	18.102	18.102
2	3.623	15.753	40.172	2.293	9.970	28.073
3	1.947	8.464	48.635	2.057	8.945	37.017
4	1.372	5.964	54.600	1.997	8.683	45.700
5	1.170	5.085	59.685	1.852	8.053	53.753
6	1.084	4.714	64.399	1.744	7.581	61.334
7	1.012	4.401	68.800	1.717	7.465	68.800

Table 2. Final results of EFA.

	F1 Responsiveness	F2 Distractors	F3 State of the technological means	F4 Initiative	F5 State of the seats	F6 State of the boards	F7 Reliability of the technological means
Cronbach α	0.88	0.71	0.73	0.745	0.88	0.82	0.76
C7	0.758	0.090	−0.128	0.084	−0.117	0.087	0.202
C8	0.686	−0.079	−0.184	0.264	−0.010	0.070	0.186
C12	0.786	0.124	0.019	0.202	−0.108	−0.007	0.032
C26	0.762	0.100	0.163	0.169	0.079	−0.003	−0.074
C27	0.786	−0.029	0.285	0.151	0.036	−0.024	−0.019
C30	0.685	0.013	0.138	0.303	0.190	0.059	0.052
C14	0.704	−0.049	0.050	0.040	−0.030	0.251	−0.084
C57	0.016	0.785	0.092	0.068	0.039	−0.059	0.245
C58	0.039	0.666	0.164	−0.069	0.081	0.161	−0.038
C59	0.035	0.636	−0.036	0.072	0.164	0.138	−0.093
C60	0.032	0.680	0.089	−0.041	0.288	0.128	0.228
C43	0.104	0.163	0.771	−0.002	0.024	0.027	0.325
C44	0.080	0.215	0.783	0.035	0.131	0.005	0.224
C38	0.044	−0.060	0.670	0.133	−0.129	0.335	−0.002
C1	0.219	−0.025	0.033	0.839	−0.054	−0.010	0.062
C2	0.319	−0.064	0.049	0.723	0.060	0.028	0.034
C16	0.349	0.160	0.073	0.683	−0.065	0.034	−0.063
C46	−0.021	0.231	0.060	0.000	0.890	0.108	0.058
C48	0.000	0.246	−0.033	−0.043	0.899	−0.016	0.091
C50	0.152	0.124	0.139	−0.006	0.061	0.859	0.118
C61	0.106	0.260	0.102	0.042	0.055	0.826	0.196
C54	0.038	0.094	0.166	0.025	0.068	0.113	0.828
C55	0.093	0.108	0.325	0.023	0.095	0.201	0.768

(3) estimation, (4) fit test, (5) interpretation, and (6) re-specification. The maximum likelihood estimation method was used to conduct the CFA because it is robust and even useful in situations of extreme non-normality (Lam & Bae, 2013; West, Finch, & Curran, 1995). CFA was also used to identify multivariate outliers through the Mahalanobis Distance (MD2) test, assuming a criterion of 0.003 as the threshold value. Thus, four observations were eliminated, achieving a final sample of 290 data. Testing the assumptions of the CFA, most of the variables showed negative asymmetries (17 of 23 variables), ranging from −0.070 to −0.982; the asymmetries for the remaining variables ranged between 0.424 and 1.755. The kurtosis coefficients ranged between −0.009 and −1.041 for 14 variables, 0.046 and 0.486 for the 7 others, and between 1.275 and 2.801 for 2 more variables.

These values are characteristic when assessing customer satisfaction and service quality, but are not necessarily a problem for CFA (Kiran & Diljit, 2012; Pallant, 2007). For example, Kline (cited in Carreras, 2006) argues this would not be a problem if for each indicator (or variable) the asymmetry indices are less than or equal to ± 3 and the kurtosis is less than 8. Also, with respect to multivariate normality, the kurtosis was 66.049 (Mardia's coefficient), with a critical ratio of 16.584. These values meet the terms reported by Rodríguez and Ruiz (2008), who argue that estimation through the maximum likelihood method provides the best results when Mardia's coefficient is less than or equal to 70, even if it does not meet the multivariate normality. This is consistent with that reported by several authors (Browne, 1984; West et al., 1995), who commented that the maximum likelihood method is robust although it does not meet the multivariate normality.

In addition to this, Hair et al. (1998) commented that large samples favour the stability of results when there is no normality; up to 200 observations (Rial et al., 2006) – ours is a 290-observation sample. The final model is tested with the generalised least-square method, which is more flexible to normality assumptions, even though the maximum likelihood estimation is the basis method (Bollen, 1989).

It should be noted that the re-specification is used to find a better match for the global model, if the empirical evidence suggests so. The model re-specification was carried out step by step and overall, considering the global fit, factor loading significance, and reliability indices of each item (deleting items with squared multiple correlations is suggested, <0.4; e.g. Franke, Von, & Schreier, 2006; Babin & Boles, 1998), as well as the modification indices. This allows comparing the original model against alternative models derived from the re-specification.

3.3.5. *Final model validity and reliability*

In order to determine the reliability and validity of the final model, the following were used: *Standardised validity coefficients* (λs), which are the loadings of each item into the respective latent factor, which are considered reasonable when their coefficients are significant, at .05 (Rial et al., 2006). However, in this study we considered the significance level to be at .01.

Convergent validity, which refers to the fact that the indicators of each construct share a high proportion of variance among them; it was calculated using the average variance extracted (AVE), accepting values of at least 50%. This was calculated using Equation (1) (Fornell & Larcker, 1981; Rial et al., 2006).

$$\mathrm{AVE}_j = \frac{\sum_{i=1}^{n_j} \lambda S_{ij}^2}{\sum_{i=1}^{n_j} \lambda S_{ij}^2 + \sum_{i=1}^{n_j} \varepsilon_{ij}}, \tag{1}$$

where 'λS' is the standardised factor loading for the 'i' indicator (or observable variable), belonging to the 'j' latent factor, and 'ε_{ij}' is the respective error term.

Discriminant validity, which is a measure of how much a specific construct differs from others, implies that each construct shares more variance with its own indicator than with other constructs (Rial et al., 2006). That expression is detailed in Equation (2).

$$\mathrm{AVE}_j > \mathrm{Cor}(\xi_j,\ \xi_k)^2\ \forall k \neq j\ , \tag{2}$$

where '$\mathrm{Cor}(\boldsymbol{\xi}_j, \boldsymbol{\xi}_k)$' denotes the inter-construct correlation coefficient.

Composite reliability, which is understood as the degree of internal consistency between indicators of the same construct, accepting values of 0.7 or higher. The expression is presented in Equation (3) (Fornell & Larcker, 1981; Rial et al., 2006).

$$\rho_j = \frac{\left(\sum_{i=1}^{n_j} \lambda S_{ij}\right)^2}{\left(\sum_{i=1}^{n_j} \lambda S_{ij}\right)^2 + \sum_{i=1}^{n_j} \varepsilon_{ij}}. \tag{3}$$

4. Results

The initial model, resulting from EFA and composed of 23 items and 7 factors, was taken to a path diagram in CFA (see Figure 2).

The results on the performance indices for the initial model are presented as follows, with acceptance criteria in parentheses: χ^2, 430.728; degrees of freedom, 209; p-value, .000; ratio (χ^2/df), 2.1 (<2); root mean square error of approximation (RMSEA), 0.061 (<0.05); goodness-of-fit index (GFI), 0.888 (>0.90); adjusted goodness-of-fit index (AGFI), 0.852 (>0.90); CFI, 0.918 (>0.95); SRMR, 0.050 (<0.05); and normed fit index (NFI), 0.855 (>0.90). It can be concluded that the initial model does not yield an adequate fit to the empirical evidence, making its re-specification required. First, such critical reasons as factor saturations are reviewed, checking that all the parameters are significant at the .01 level. And, precisely, all were found to be significant at .001 (see Appendix 1). Second, underscoring the reliability of each of the indicators, some are found to not reach the minimum value of 0.4 (Babin & Boles, 1998; Franke et al., 2006): C38 (0.198), C59 (0.235), C58 (0.266), and C14 (0.391), which initially raised the sequential removal of these items. On the other hand, the modification indices reported cross-saturations for most of these items (loaded significantly on other factors), while their terms of error share meaningful relationships with other items of the model. Once each of the re-specifications of the model was executed sequentially, and considering each change for both its reliability and cross-saturation indices, a model composed of 7 latent dimensions, 17 indicators, and 98 degrees of freedom was reached (see Figure 3).

The results obtained for model 1 were: χ^2, 144.007; p-value, .002; degrees of freedom, 98; ratio (χ^2/df), 1.469 (<2); RMSEA, 0.040 (<0.05); GFI, 0.947 (>0.90); AGFI, 0.917 (>0.90); CFI, 0.977 (>0.95); SRMR, 0.036 (<0.05); and NFI, 0.933 (>0.90). This model achieves a goodness of fit index significantly better than the initial model though the χ^2 value remains significant due to the large size of the sample. In order to explore opportunities for improvement of model 1, another re-specification is planned. For this purpose, the C30 item was removed since it presented cross-saturations on other factors (chairs) and because its error terms shared meaningful relationships with several variables of the model. Moreover, taking into account the semantic content of the responsiveness

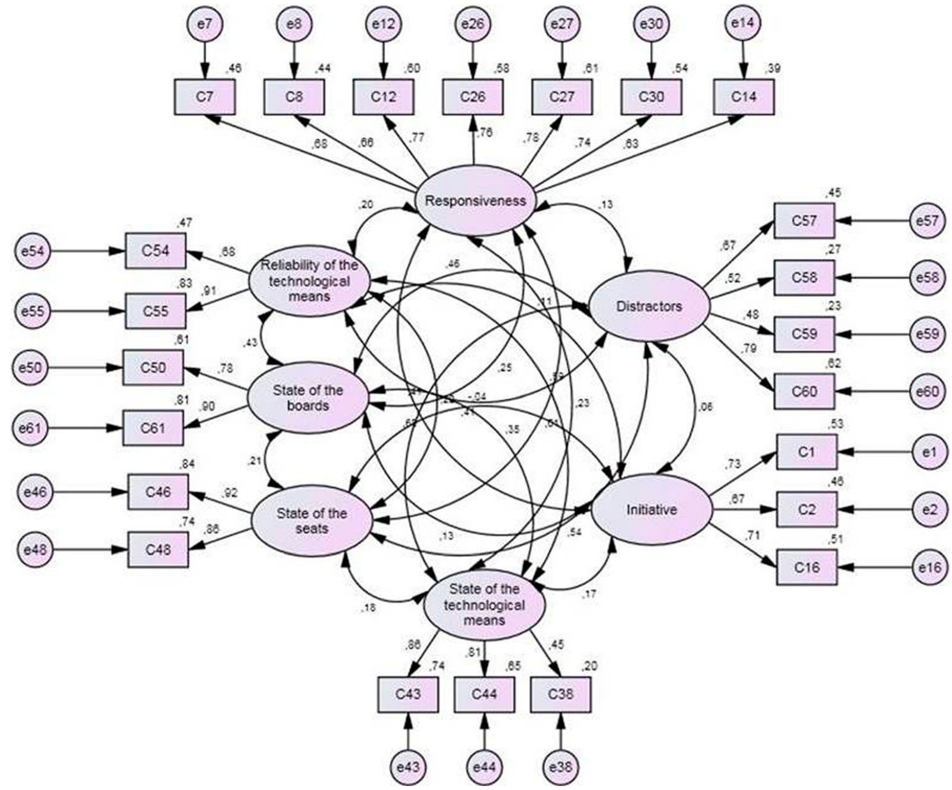

Figure 2. Initial model (composed of 7 factors, 23 indicators, and 209 degrees of freedom).

factor, it was considered that upon excluding the C30 item, the other items represent its content. The resulting pattern of this re-specification has been named model 2 (see Figure 4).

Next, the fit indices of model 2 are presented: χ^2, 110.774; *p*-value, .023; degrees of freedom, 83; ratio (χ^2/df), 1.334 (<2); RMSEA, 0.034 (<0.05); GFI, 0.955 (>0.90); AGFI, 0.927 (>0.90); CFI, 0.984 (>0.95); SRMR, 0.0325 (<0.05); and NFI, 0.942 (>0.90).

It is worth noting that the overall fit of model 2 is better than that of model 1 and consequently better than that of the initial model. Moreover, all items showed standardised validity coefficients that met the criteria of significance assumed; Rial et al. (2006), for instance, suggests a .05 significance level; all loads in model 2 were significant at .001. Similarly, Franke et al. (2006) and Babin and Boles (1998) suggest standardised factor loadings at a minimum of 0.5; model 2 meets this criterion in all cases, as the minimum load was 0.62 (C57). Regarding the reliability of each item, the squared multiple correlations were higher than 0.4, except for C57 (0.38); however, it was decided that it be kept in the model because its value was close to the threshold and presented factor loadings that satisfied the criterion validity of the item. Also, in conjunction with C60, C57 fully met the criteria of reliability of the complete factor (as shown below).

In order to explore other possible scaled models and considering possibilities for improvement of model 3 from the modification indices and the semantic nature of the items, so that all would be represented, two more models were reached. Model 3

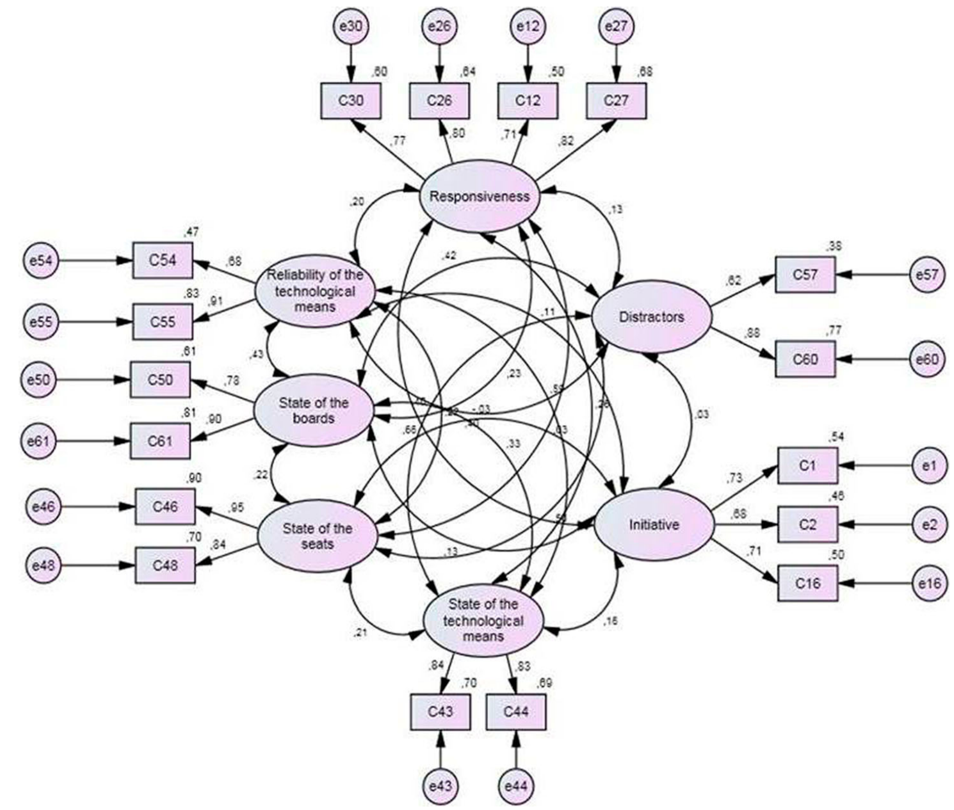

Figure 3. Model 2 (composed of 7 factors, 17 indicators, and 98 degrees of freedom).

corresponds exactly to model 2, with the exclusion of item 12 of factor 1 (responsiveness), for a total of 15 items. In parallel, model 4, corresponds to model 3, excluding item 16 of factor 3 (initiative), for a total of 14 items; that is, in model 4 each dimension had the minimum number of items from a practical point of view (2 items, e.g. Rial et al., 2006). It should be noted that, from the theoretical point of view, this last model does not present identification problems because its degrees of freedom are greater than zero (56). Also, the errors are assumed to be uncorrelated, which is assumed in the most confirmatory models (Rial et al., 2006). However, this does not exclude the fact that the model can be empirically unidentified (Kenny, 1979), demanding the increase in the sample size, modifying the model, or setting a value to problematic parameters (Rial et al., 2006).

For the five models in Table 3, the results of overall fit to the empirical evidence are summarised and reliability indices of total scales (Cronbach α) are added, as well as the results of convergent validity and present composite reliability factors.

Considering the fit indices of the global model (χ^2 ratio/df, RMSEA, GFI, AGFI, CFI, SRMR, and NFI), note that models 1–4 have a satisfactory fit. When taking into account χ^2 (p-value), model 1 does not satisfy it (p-value: .002), neither does model 2, though it is close (.023), while models 3–4 satisfy it (p-value $>$.05). However, the χ^2 test is sensitive to sample size and complexity of the model. Therefore, the literature reflects consensus on the need to compare their results with other indices of overall fit and also take them together with the degrees of freedom of the model, using the ratio χ^2/df (Hu & Bentler,

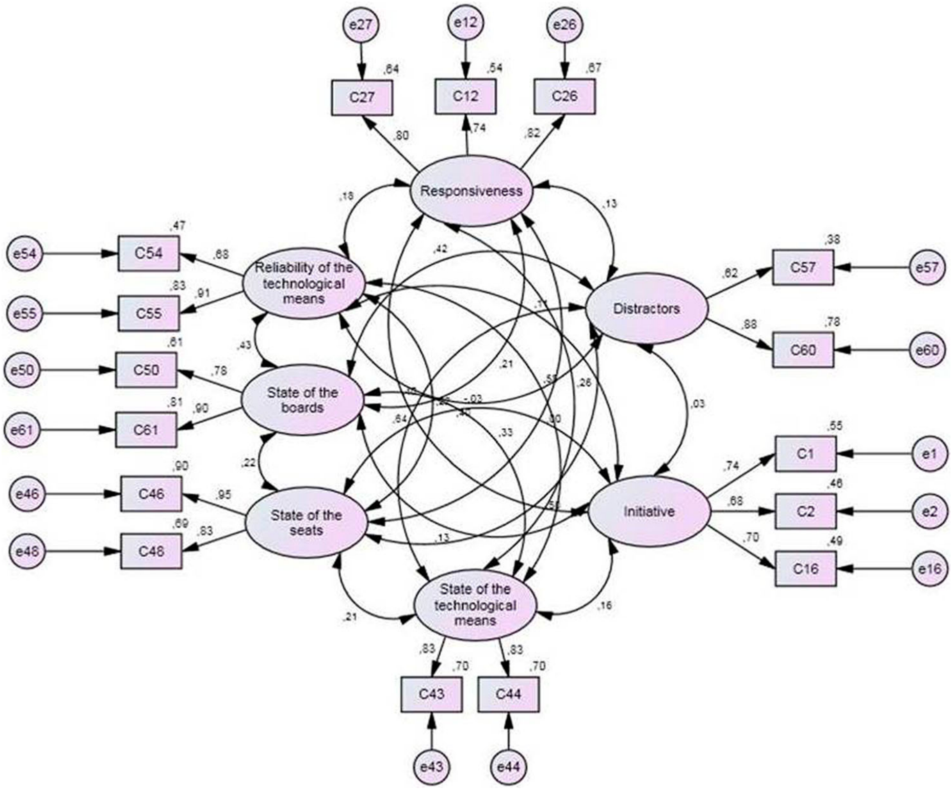

Figure 4. Model 3 (composed of 7 factors, 16 indicators, and 83 degrees of freedom).

1995; Jöreskog & Sörbom, 1993; Rial et al., 2006). Taking into account the results of global fit (Table 3), it can be concluded that models 1–4 reflect an adequate fit with the empirical evidence. Also, when the composite reliability and validity of the models' convergent measurement are taken into account (using Equations (1) and (3)), results are favourable. Note that the factors of the four models satisfy the criterion of minimum 0.5 in AVE (convergent validity). As for the composite reliability factors, models 1–3 meet the minimum criteria of 0.7, but factor 3 (initiative) of model 4 does not satisfy them (0.68).

Contrasting these models in light of parsimony (AIC) and adjustability with other hypothetical samples of the same size (ECVI), useful measures to the comparison of theoretical models, models with lower values of these indices are preferred (Jöreskog & Sörbom, 1993; Rial et al., 2006). So, regarding parsimony indices, better results are obtained for model 4 (AIC: 164.20; ECVI: 0.57), followed by model 3 (AIC: 189.97; ECVI: 0.66), model 2 (AIC: 216.77; ECVI: 0.75), and, finally, model 1 (AIC: 254.1; ECVI: 0.88).

Focusing, then, on the results of global fit, convergent validity, composite reliability, parsimony, and replicability with other samples, as well as the semantic content of the items in each factor, it is suggested to proceed with models 2 and 3. Model 1 is discarded because, while it is set globally for empirical evidence (as it also occurred with 2–4 models), excluding only one item from factor 1 (leading to model 2), maximum two

Table 3. Summary of fit, composite reliability, and convergent validity of the initial model and its four alternative models.

Indexes	Accepted value	Initial model	Model 1	Model 2 (Model 1 without C30)	Model 3 (Model 2 without C12)	Model 4 (Model 3 without C16)
Ítems	N/A	23	17	16	15	14
χ^2	Lower	430.728	144.007	110.774	87.974	66.202
p-Value	<0.05	0.000	0.002	0.023	0.061	0.165
df	>0	209	98	83	69	56
Ratio χ^2/df	<2	2.1	1.469	1.334	1.275	1.182
RMSEA	<0.05	0.061	0.040	0.034	0.031	0.025
GFI	>0.90	0.888	0.947	0.955	0.963	0.969
AGFI	>0.90	0.852	0.917	0.927	0.935	0.943
CFI	>0.95	0.918	0.977	0.984	0.988	0.993
SRMR	<0.05	0.050	0.036	0.032	0.031	0.028
NFI	>0.90	0.855	0.933	0.942	0.949	0.958
AIC	Lower	564.728	254.077	216.774	189.974	164.202
ECVI	Lower	1.954	0.879	0.750	0.657	0.568
Cronbach α	>0.7	0.845	0.816	0.800	0.790	0.785

Factors	AVE	C.R.	AVE	C.R.	AVE	C.R.	AVE	C.R.
F1. Responsiveness	0.61	0.86	0.61	0.83	0.66	0.80	0.66	0.80
F2. Distractors	0.58	0.73	0.58	0.73	0.58	0.73	0.58	0.73
F3. Initiative	0.50	0.75	0.50	0.75	0.50	0.75	0.50	0.68
F4. State of the technological means	0.70	0.82	0.70	0.82	0.70	0.82	0.70	0.82
F5. State of the seats	0.80	0.89	0.80	0.89	0.80	0.89	0.80	0.89
F6. State of the boards	0.71	0.83	0.71	0.83	0.71	0.83	0.71	0.83
F7. Reliability of the technological means	0.65	0.78	0.65	0.78	0.65	0.78	0.65	0.78

items (model 3), the said factor continues to show composite reliability indices greater than 0.7, and satisfactory convergent validity (AVE >0.5). With these re-specifications to model 1, the resulting models (2 and 3) substantially improve the overall fit (χ^2 ratio/df and other indicators). Also, the parsimony and replication rates in other samples of equal size make model 1 the less appropriate model, set against the other comparison models. Moreover, as an indirect result, it should be noted that model 2 (16 items) and model 3 (15 items) are favoured over model 1 (17 items) regarding their multivariate normality, since Mardia's coefficient decreased from 66,049 (model 1) to 30,362 (model 2), and to 28,029 (model 3).

Additionally, for both models (2 and 3) the possibility that there were problems of common method variance (CMV) was examined *post hoc*, using the Harman test of a single factor by CFA. Thus, the fitted models (2 and 3) were compared against a model in which all items saturate on one latent factor (factor of engagement). In case there were problems of CMV, the model of a single factor should fit the empirical evidence better (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). The results for the model of a single factor did not even moderately fit to the data. For example, a single factor model with the 16 items of model 2 showed a χ^2: 1273.2; p-value: .000; degrees of freedom: 104; ratio (χ^2/df): 12.24 (well above the maximum accepted for 2). In a similar way, a model with the 15 items of model 3 showed: χ^2: 1108.7; p-value: .000; degrees of freedom: 90; ratio (χ^2/df): 12.31. This suggests that the variance of the items does not appear to be due to the method of evaluation.

Continuing with the process of validation of models 2 and 3, we proceeded with the analysis of discriminant validity (see Equation (2)). It is worth remembering that in order for such validity to exist, the AVE should be greater than the square of the inter-factor coefficients. To facilitate comparison between these two indicators, the AVE/Max (SIC) ratio was also calculated, understanding SIC as squared inter-construct correlations, so while this reason exceeds the value one (1), there is more evidence of discriminant validity factors. If the ratio is very close to one (1) or less than 1, there is less evidence that factors differ from each other, so it is suggested that these be united, or else that a second-order factor that contains both be studied.

Looking at Table 4, it can be seen that both models (2 and 3) have discriminant validity. It should be noted that the ratio AVE/Max (SIC), in all cases, considerably exceeded the value of 1, showing the lowest value in factor 1 (responsiveness), with a ratio of 1.5;

Table 4. Discriminant validity of the factors for models 2 and 3.

	F1		F2	F3	F4	F5	F6	F7
Factors	Model 2	Model 3	Models 2–3	Models 2–3	Models 2–3	Models 2–3	Models 2–3	Models 2–3
F1. Responsiveness	0.61	0.66						
F2. Distractors	0.02	0.02	0.58					
F3. Initiative	0.41	0.36	0.00	0.58				
F4. State of the technological means	0.07	0.08	0.16	0.03	0.7			
F5. State of the seats	0.00	0.00	0.25	0.00	0.04	0.80		
F6. State of the boards	0.04	0.04	0.17	0.02	0.11	0.047	0.71	
F7. Reliability of the technological means	0.03	0.04	0.16	0.01	0.34	0.048	0.18	0.65
AVE/Max (SIC)	1.5	1.8	2.3	21.6	2.0	16.8	3.8	

Table 5. Implementation of ClassroomQual to the service context.

Factors	Items: Version 1 (Model 2)	Items: Version 2 (Model 3)	Items loadings: Version 1 (Model 2)	Items loadings: Version 2 (Model 3)	Items average: Version 1 (Model 2)	Items average: Version 2 (Model 3)	Factor level scores: Version 1 (Model 2)	Factor level scores: Version 2 (Model 3)
F1. Responsiveness	C12		0.736		4.748		4.67	4.64
	C27		0.820	0.848	4.590			
	C26		0.799	0.787	4.693			
F2. Distractors	C57		0.618	0.617	3.248		3.07	3.07
	C60		0.880	0.881	2.948			
F3. Initiative	C1		0.740	0.741	4.959		4.87	4.74
	C2		0.676	0.682	4.783			
	C16		0.701	0.694	4.479			
F4. State of the technological means	C43		0.834	0.832	4.686		4.79	4.79
	C44		0.834	0.836	4.890			
F5. State of the seats	C46		0.951	0.950	2.317		2.19	2.19
	C48		0.832	0.833	2.045			
F6. State of the boards	C50		0.779	0.779	4.072		4.04	4.04
	C61		0.901	0.901	4.003			
F7. Reliability of the technological means	C54		0.682	0.683	4.410		4.30	4.30
	C55		0.910	0.910	4.224			

that is, the AVE in that factor (0.61) was 50% above the highest SIC of that factor (0.41). For the other factors, the results were consolidated in a cell, in each case, since the values did not suffer any changes.

Finally, comparing models 2 (16 items) and 3 (15 items), it is noted that item C12 does not significantly harm the results, either as a global fit to empirical evidence (which actually makes it better) or as the reliability of the total scale (Cronbach α) is reduced from 0.80 to 0.79 (see Table 4). Nor are the results of the factors significantly affected. For example, the composite reliability goes from 0.83 to 0.80 for the factor in which the item (F1) is removed, remaining the same for the rest, as expected. Even with respect to the convergent validity (AVE), better results are obtained, as it surpasses 0.61 (model 2) to 0.63 (model 3). Similarly, although satisfactory in both models, discriminant validity is also favoured, as in this factor (F1) the ratio AVE/Max (SIC) goes from 1.5 (model 2) to 1.8 (model 3), differing even more from the factor with greater SIC (F3 initiative). Therefore, model 3 with 7 factors and 15 items is recommended.

However, in order to encourage further studies and replication with other samples, contexts, etc., the items of model 2 are described in detail, in Annexes 2 (Spanish) and 3 (English); that is, the 16 items. It is known that excluding item 12 model 3 is reached. Thus, readers and future researchers interested in using the model can compare, choose, and release their improvements and preferences regarding this instrument, called 'ClassroomQual'. Therefore, the instrument can be used in two versions: version 1 (16 items: model 2) and version 2 (15 items: model 3). It is worth noting that, after applying the proposed scale, it is necessary to consider that the score (encoding) may differ depending on how the items are written. For example, for the positively worded items, the score ranged from (1) strongly disagree to (7) strongly agree, but for the negatively worded items, the score ranged from (1) strongly agree to (7) strongly disagree.

On the other hand, the results of applying models 2 and 3 to the context of this study are presented in Table 5. The scoring method used for each factor was performed through Fornell's Solution (Rial et al., 2006), which estimates the model parameters based on the weighted average of observable variables (indicators), taking CFA factor loadings as weighting elements.

It is worth remembering that the results obtained from models 2 and 3 are consistent, because there was only a slight variation in factor 1 (responsiveness: 4.67 model 2; 4.64 model 3) and factor 2 (initiative: 4.87 model 2; 4.74 model 3). It is important to keep in mind that model 3 does the same with one item less. Appendices 2 and 3 show the ClassroomQual instrument in Spanish and English, respectively.

5. Discussion

The results obtained complement contributions by Miller and Brooks (2010) regarding class service quality (ClassQual). These authors provide elements to understand such service quality, emphasising teaching, which is considered the core service within the typologies proposed by Grönroos (1992) and Nóbrega (1997). However, within the framework of higher education service, complementary and supplementary (or facilitating and support) services should also be considered for a better understanding of service quality in higher education. This is because, currently, there are few studies on these service types, and improvement to these may give a competitive advantage to institutions (Chen et al., 2013; Miller & Brooks, 2010).

In this paper, the use-of-classrooms-for-teaching–learning service is shown as a service complementary to the core service, which generates needs and expectations in students. In the specific context of this study, there are seven dimensions: responsiveness, initiative, state of technological means, reliability of technological means, state of whiteboards, state of seats, and distractors. These lead us to differentiate the support staff in terms of technical aspects and behaviour. In addition, this is consistent with two of the SERVQUAL dimensions (reliability and responsiveness) and with non-academic aspects of HedPERF.

On the other hand, the 'state of technological means' (F4) and 'reliability of technological means' (F5) dimensions are consistent with University of Antioquia classrooms, which for the most part have computers and projectors. However, these dimensions are related to tangibles and reliability dimensions of SERVQUAL, which are not adequately represented by that instrument. In fact, Chen et al. (2013) argue that SERVQUAL is more appropriate for high personal interaction services than for services of high technical quality. For example, note that the 'reliability' dimension of ClassroomQual, unlike SERVQUAL, is more specific with regard to the teaching means (e.g. computer equipment) performing well through time. Surprisingly, ClassroomQual distinguishes between two moments of the service. One of the moments is before starting class, in which the teaching means (PC and projector) may be damaged (state of technological means), and the other moment refers to the performance of these during class (reliability). A possible explanation for this, in the observation context, is that failures of the means of teaching during these two moments may have different consequences for students. If the computer or projector is damaged before starting class, the professor can have the equipment or the classroom changed through support personnel.

However, if an equipment fails during class, the consequences could perhaps be more critical, since this not only causes idle time, but also interrupts the professor's lecture and, might hinder students' attention after the problem has been solved.

In this study, factors related to the teaching environment were identified: state of seats and distractors. Once again, this result seems to be consistent with the context of the study, because, in the classrooms, the seats were sometimes broken, and this might cause inconvenience to the students during class. Similarly, in the public universities context, it is possible that a large number of students assigned to a class (in excess of 20–25 students) held in too small a classroom inconvenience students and affect the teaching–learning levels for them (core service). In summary, the core service (teaching–learning) may be facilitated (delivered) on three dimensions: (1) attention to needs of the professor at the beginning, during, or at the end of a class; (2) the teaching means used by the professor to transmit information to students; and (3) the teaching environment in which the class is being conducted.

The complementary service identified in this study, named 'use-of-classrooms-for-teaching–learning', together with the results found by Miller and Brooks (2010) regarding course content dimensions and professor elements, can lead to a better understanding of the many factors that influence the satisfaction of students' needs and expectations. Similarly, in an overview of service quality from the students' perspective, the results of this study may complement some of the dimensions of HedPERF and SERVQUAL/SERVPERF, among others, in noteworthy aspects related to such departments as administrative, library, student welfare, etc.

The results obtained also present similarities with the other reviewed instruments. The HEDQUAL, for instance, has two items, 'rapid service' and 'knowledge about systems and procedures', grouped into the quality dimension, 'administrative services quality'.

Furthermore, the dimension termed 'support services quality' contains two items: 'size of the classroom' and 'necessary equipment in the classrooms (computer, digital projector, etc.)'. The ClassroomQual scale expands these elements further, differentiating, for instance, between the state/availability of informatics equipment and their reliability. Likewise, ClassroomQual shows the importance of evaluating the 'initiative' factor, in addition to the service speed.

On the other hand, the HiEdQUAL scale also encompasses aspects regarding support personnel and classrooms. In particular, the 'administrative services' dimension contains three items regarding complementary services: 'administrative staff provides service without delay', 'administrative staff is courteous and willing to help', and 'administrative staff provides error free work'. Moreover, the dimension 'academic facilities' contains another three items: 'department contains adequate facilities', 'classrooms are equipped with teaching aids', and 'department has sufficient classrooms'. However, due to its generalist scope, the HiEdQUAL does not explore in depth 'during class' experience details, such as state of boards, state of the chairs, and 'distractors', all of them featured in the ClassroomQual scale. Additionally, the HiEdQUAL does not differentiate between the state and reliability of the informatics equipment.

Finally, it is worth mentioning the study developed by Stukalina (2014). This author proposes two indicators for predicting student satisfaction and motivation regarding the higher education service: the environment comfort where the service is performed and the adequate assistance given by the support staff. Nevertheless, that study deals with these indicators in a general form, being impossible to identify specific service quality attributes inherent to them.

The present results reinforce the relevance of expanding the complementary services, from the student perspective, to achieving improved quality in higher education. High standards in parallel services might contribute to higher standards of the core service performance: teaching–learning. Likewise, the study remarks on the need to develop new measurement instruments that allow the identification and assessment of the different complexities underlying each service, in order to supplement the generalist perspectives within each context's details and customer typologies.

Also, the results suggest relevant managerial implications for complementary services contexts similar to University of Antioquia's. Particularly, in this institution, a department that is independent of the university academic units offers the service. This department is also the provider of other complementary services such as logistic operations and informatics services, among others. Moreover, the complementary service studied is the 'core service' of that department and its value-offering might be strongly supported by results. Concerning the application of the results, service quality level from the students' viewpoint is worrisome. The scores for service quality dimensions ranged from 2.19 to 4.79 (with a mean of 3.99), leading to the understanding that service performance, as assessed by students, tends to be unfavourable (1 = very poor quality; 7 = very good quality). After normalising the score to understand it better (0–100%), it can be seen that service performance level is only 49.8%. This, perhaps, may be harmful to the core service (teaching–learning).

6. Conclusions

In the context of the present study, the multidimensionality of service quality was confirmed. The initial test model in CFA consisted of 23 items, derived from the exploratory phase, and grouped into 7 latent factors. Through strategies of re-specification, 4 models were reached, and finally 2 models (each with 7 factors) are proposed as the top performers in terms of overall fit, reliability, and validity. One model contains 16 items and the other 15, and there are two versions of the ClassroomQual scale. This presents a structure with global fit, parsimony, and satisfactory reliability and validity. ClassroomQual provides the elements for a better understanding of the structure of complementary-service quality and makes further analysis and replications possible. The novelty of this work lies in presenting the structure of a service barely studied in the context of higher education, as a new measurement instrument. The ClassroomQual instrument aims to motivate other researchers to study complementary services, which can be as important as core services, that is, teaching–learning, in this case.

Similarly, understanding the structures of the service quality studied might enhance future research on the effects of the service quality and their components on essential service dimensions. ClassroomQual is a scale that provides elements for a better understanding of the complementary service studied, and it leads to several opportunities for research about the service incidences on other constructs (intrinsic to students and essential services, among others). This paper also makes it possible to carry out further studies for the improvement of ClassroomQual.

Regarding managerial implications, the present findings emerge as useful guidelines to aid and improve decision-making processes in educational institutions, concerning resource planning, among others. The 'teaching means', among other aspects, plays a vital role in the student's perception of quality regarding the use-of-classroom service. It provides evidence of the increasing importance of ICT to the current education paradigm, where the globalisation of knowledge and informational literacy are key to the teaching–learning process. Moreover, 'teaching environment' highlights the importance of having students enjoy an appropriate environment during the teaching–learning process. Finally, 'support personnel' demonstrates, once more, the importance of the human performance component to service delivery, in order to achieve higher levels of quality. Therefore, institutions must pay special attention to their managerial practices on these three dimensions of service delivery, besides the well-known professor competencies, in order to better overall education quality.

7. Limitations and future research

The measurement of the use-of-classrooms-for-teaching–learning service quality, from the student perspective, involved students from the industrial engineering programme solely, limiting the generalisation of the results. In this sense, future replications of the study encompassing student perceptions from different programmes would allow for the validity of the scale to be confirmed. Likewise, it must be considered that the complementary services assisting the teaching–learning process might vary in form and content among different higher education institutions. Moreover, they might vary among different departments or faculties within the same institution. Therefore, future studies should be replicated in different faculties as well as different higher education institutions, in order to explore and understand better the nature of complementary services.

In parallel, involving the perceptions of other customer types (professors, administrative personnel, etc.) would help with understanding the underlying complexity regarding this service typology. At the methodological level, the design of future longitudinal and experimental studies would make it possible to understand the dynamics underlying the studied service, also helping identify the causes and effects of customer perceptions. Also, it is important to highlight that the survey was administered in Spanish and, therefore, the English version has not been validated.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Jorge Pérez Rave http://orcid.org/0000-0003-1166-5545

References

Abdullah, F. (2006a). Measuring service quality in higher education: HEdPERF versus SERVPERF. *Marketing Intelligence & Planning*, *24*(1), 31–47.
Abdullah, F. (2006b). The development of HEdPERF: A new measuring instrument of service quality for the higher education sector. *International Journal of Consumer Studies*, *30*(6), 569–581.
Annamdevula, S., & Bellamkonda, R. (2012). Development of HiEdQUAL for measuring service quality in Indian higher education sector. *International Journal of Innovation, Management and Technology*, *3*(4), 412–416.
Astin, A. (1993). *What matters in college? Four critical years revisited*. San Francisco, CA: Jossey-Bass.
Babin, B., & Boles, J. (1998). Employee behavior in a service environment: A model and test of potential differences between men and women. *Journal of Marketing*, *62*(2), 77–91.
Barani, G., & Kumar, R. (2013). The impact of service quality and scholars delightfulness of ASHE in private universities of Tamilnadu, India. *Life Science Journal*, *10*(1), 2801–2809.
Bollen, K. (1989). *Structural equations with latent variables*. New York, NY: John Wiley and Sons.
Browne, M. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62–83.
Carreras, E. (2006). El pronóstico en modelos estructurales: predecir la satisfacción latente. En J. Lévy & J. Varela (Dir. y Comp.). *Modelización con estructuras de covarianzas en ciencias sociales* (pp. 453–488). La Coruña, España: Editorial Netbiblo S.L. (Editora: Seco, C).
Chen, N., Huang, S., Shu, S., & Wang, T. (2013). Market segmentation, service quality, and overall satisfaction: Self-organizing map and structural equation modeling methods. *Quality & Quantity*, *47*(2), 969–987.
Chua, C. (2004). Perception of quality in higher education. In R. Carmichael (Ed.), *Quality in a time of change*, Proceedings of the Australian Universities Quality Forum (Vol. 4, pp. 181–187). Melbourne: AUQA Occasional Publication.
Cronin, J., & Taylor, S. (1992). Measuring service quality: Re-examination and extension. *Journal of Marketing*, *56*(3), 56–68.
Dlačića, J., Arslanagićb, M., Kadić-Maglajlićb, S., Markovićc, S., & Raspord, S. (2014). Exploring perceived service quality, perceived value, and repurchase intention in higher education using structural equation modelling. *Total Quality Management & Business Excellence*, *25*(1–2), 141–157.
Firdaus, A. (2006). Measuring service quality in higher education: Three instruments compared. *International Journal of Research & Method in Education*, *29*(1), 71–89.
Fornell, C., & Larcker, D. (1981). Evaluating structural equation models with unobservable variables and measurement errors. *Journal of Marketing Research*, *18*, 39–50.
Franke, N., Von, H. E., & Schreier, M. (2006). Finding commercially attractive user innovations: A test of lead-user theory. *Journal of Product Innovation Management*, *23*(4), 301–315.
Godino, J. (2011, May 13). *Indicadores de la idoneidad didáctica de procesos de enseñanza y aprendizaje de las matemáticas*. Conferência Interamericana de Educação matemática-ciaem, Recife, Brasil.

Grönroos, C. (1992). *Service management and marketing: Managing the moments of truth inservice competition*. New York, NY: Maxwell Macmillan International Student Editions.
Hair, J., Anderson, R., & Tatham, R. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.
Hayes, B. (2002). *Cómo medir la satisfacción del cliente. Desarrollo y utilización de cuestionarios* (3ª ed.). Barcelona: Gestión 2000.
Hittman, J. (1993). TQM and CQI in post secondary education. *Quality Progress*, *26*, 77–80.
Hu, L.-T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Thousand Oaks, CA: Sage.
Icli, G., & Anil, N. (2014). The HEDQUAL scale: A new measurement scale of service quality for MBA programs in higher education. *South African Journal of Business Management*, *45*(3), 31–43.
Jiewanto, A., Laurens, C., & Nelloh, L. (2012). Influence of service quality, university image, and student satisfaction toward WOM intention: A case study on Universitas Pelita Harapan Surabaya. *Procedia-Social and Behavioral Sciences*, *40*, 16–23.
Jöreskog, K., & Sörbom, D. (1993). *LISREL 8 user's reference guide*. Chicago: Scientific Software International.
Kadhim, R., Taqi, B., & Shuaibu, B. (2012). Development a prototype of academic performance among university students. *International Journal of Independent Research Studies*, *1*(1), 39–49.
Kenny, D. A. (1979). *Correlation and causality*. New York, NY: Wiley-Interscience.
Kessler, S. (1995). *Total quality service: A simplified approach to using the Baldrige award criteria*. Milwaukee: ASQC Quality Press.
Kiran, K., & Diljit, S. (2012). Modeling web-based library service quality. *Library & Information Science Research*, *34*, 184–196.
Koni, A., Zainal, K., & Ibrahim, M. (2013). An assessment of the services quality of Palestine Higher Education. *International Education Studies*, *6*(2), 33–48.
Lam, E., & Bae, J. (In press). Confirmatory factor análisis of the purchaser style inventory for sport products (PSISP). *Sport Management Review*.
Law, D. (2013). Initial assessment of two questionnaires for measuring service quality in the Hong Kong post-secondary education context. *Quality Assurance in Education*, *21*(3), 231–246.
Marsh, H. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, *11*(3), 253–388.
Marshall, G., & Murdoch, I. (2001). Service quality in the marketing of consulting engineers. *International Journal of Construction Marketing*, *3*, 534–559.
Mattick, R., & Clarke, J. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy*, *36*(4), 455–470.
Miller, R., & Brooks, N. (2010). Class service quality: Moving beyond SERVQUAL. *Proceedings of the information systems educators conference (ISECON)* (Vol. 27, No. 1394, pp. 1–8), Nashville, TN.
Mohamad, A., Hassan, Z., Rahman, S., & Ghouri, A. (2012). Educational service quality at public higher educational institutions: A proposed framework and importance of the sub-dimensions. *International Journal of Economics Business and Management Studies*, *1*(2), 36–49.
Muñoz, L., & Pérez, J. (2013). Calidad del servicio de uso de aulas para la enseñanza-aprendizaje desde la perspectiva del estudiante. *Universidad, ciencia y Tecnología*, *17*(69), 161–169.
Nóbrega, K. (1997). *Gestão da qualidade em serviços* (Doctoral dissertation). Escola Politécnica - Departamento de Engenharia de Produção, Universidade de São Paulo, São Paulo, Brazil.
Nóbrega, K., Mendonca, C., De Souza, T., & Campos, D. (2010). *The importance of central, complementary and supplementary services offered by a university in students' retention* (Working Paper). Natal – RN: Universidade Potiguar.
Nunnally, J. (1987). *Teoría Psicométrica*. Méjico: Trillas.
Pallant, J. (2007). *SPSS survival manual: A step-by-step guide to data analysis using SPSS for windows* (3rd ed.). Sydney: Allen & Unwin.
Parasuraman, A., Berry, L., & Zeithaml, V. (1991). Refinement and reassessment of the SERVQUAL scale. *Journal of Retailing*, *67*(4), 420–450.
Parasuraman, A., Zeithaml, V., & Berry, L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, *64*(1), 12–40.

Pérez, J., & Muñoz, L. (2014). Espacio literario relevante sobre la evaluación de la calidad del servicio: países de realización de los estudios, índices de fiabilidad, hipótesis y desafíos. *Ingeniería Investigación y Tecnología, UNAM*, *XV*(3), 479–492.

Podsakoff, P., MacKenzie, S., Lee, J., & Podsakoff, N. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*, 879–903.

Quinn, A., Lemay, G., Larsen, P., & Johnson, D. (2009). Service quality in higher education. *Total Quality Management & Business Excellence*, *20*(2), 139–152.

Rasli, A., Shekarchizadeh, A., & Iqbal, M. (2012). Perception of service quality in higher education: Perspective of Iranian students of Malaysian universities. *International Journal of Economics and Management*, *6*(2), 201–220.

Reeves, C., & Bednar, D. (1994). Defining quality: Alternatives and implications. *The Academy of Management Review*, *19*(3), 419–445.

Rial, A., Valera, J., Abalo, J., & Lévy, J. (2006). El análisis factorial confirmatorio. En: Lévy, J., &Varela, J. (Dir. y Comp.), *Modelización con estructuras de covarianzas en ciencias sociales* (pp. 119–143). La Coruña, España: Editorial Netbiblo S.L. (Editora: Seco, C).

de Rijke, P., & Plucker, J. A. (2011). Getting what we wish for: The realities of business education for a global economy. *Business Horizons*, *54*, 375–382.

Rodríguez, M., & Ruiz, M. (2008). Atenuación de la asimetría y de la curtosis de las puntuaciones observadas mediante transformaciones de variables: Incidencia sobre la estructura factorial. *Psicológica*, *29*(2), 205–227.

Schneider, B., Hanges, P., Goldstein, H., & Braverman, E. (1994). Do customer service perceptions generalize? The case of student and chair ratings of faculty effectiveness. *Journal of Applied Psychology*, *79*(5), 685–690.

Shank, M., Walker, M., & Hayes, T. (1995). Understanding professional service expectation: Do we know what our students expect in quality education? *Journal of Professional Services Marketing*, *13*(1), 71–89.

Sheeja, V., Krishnaraj, R., & Harindranath, R. (2014). Evaluation of higher education service quality scale in pharmaceutical education. *International Journal of Pharmaceutical Sciences Review & Research*, *28*(2), 263–266.

Sorayael, A., Ahangar, R., Kasiri, E., & Baharami, A. (2013). An investigation into the determinant of service quality among management students (empirical study of Iranian students). *Research Journal of Applied Sciences, Engineering and Technology*, *5*(16), 1883–1891.

Stodnick, M., & Rogers, P. (2008). Using SERVQUAL to measure the quality of the classroom experience. *Decision Sciences Journal of Innovative Education*, *6*(1), 115–132.

Stukalina, Y. (2014). Identifying predictors of student satisfaction and student motivation in the framework of assuring quality in the delivery of higher education services. *Business, Management and Education*, *12*(1), 127–137.

Temtime, Z., & Mmereki, R. (2011). Challenges faced by graduate business education in Southern Africa: Perceptions of MBA participants. *Quality Assurance in Education*, *19*(2), 110–129.

Thompson, K., Haddad, L., & Smith, S. (2014). Reliability and validity of the postepidural fall risk assessment score. *Journal of Nursing Care Quality*, *29*(3), 263–268.

West, S., Finch, J., & Curran, P. (1995). Structural equation models with non-normal variables. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 56–75). Thousand Oaks, CA: Sage.

Yeo, R., & Li, J. (2012). Beyond SERVQUAL: The competitive forces of higher education in Singapore. *Total Quality Management & Business Excellence*, *25*(1–2), 95–123.

Yildiz, S. M. (2012). Instruments for measuring service quality in sport and physical activity services. *Collegium Antropologicum*, *36*(2), 689–696.

Yildiz, S. (2014). Service quality evaluation in the school of physical education and sports: An empirical investigation of students' perceptions. *Total Quality Management*, *25*(1), 80–94.

Zhao, X. (2012). A review on service quality and student satisfaction of higher education. *Soft Computing in Information Communication Technology, Advances in Soft Computing*, *158*, 115–122.

Zulkefli, N., & Uden, L. (2013). A service quality framework for higher education from the perspective of service dominant logic. *Advances in Intelligent Systems and Computing*, *172*, 307–317.

Appendix 1. Regression weights for the initial model

			Estimate	S.E.	C.R.	*p*	Label
C7	←	Responsiveness	1.000				
C8	←	Responsiveness	0.959	.092	10.392	***	par_1
C12	←	Responsiveness	1.023	.087	11.783	***	par_2
C26	←	Responsiveness	0.968	.087	11.074	***	par_3
C27	←	Responsiveness	1.038	.092	11.319	***	par_4
C30	←	Responsiveness	1.018	.094	10.788	***	par_5
C38	←	StatMeans	0.487	.068	7.154	***	par_6
C44	←	StatMeans	0.801	.067	12.050	***	par_7
C43	←	StatMeans	1.000				
C57	←	Distrac	1.000				
C58	←	Distrac	0.704	.096	7.347	***	par_8
C59	←	Distrac	0.602	.088	6.801	***	par_9
C60	←	Distrac	1.060	.110	9.647	***	par_10
C1	←	Iniciat	1.000				
C2	←	Initiat	0.967	.102	9.496	***	par_11
C16	←	Initiat	0.963	.099	9.763	***	par_12
C55	←	ReliabMeans	1.000				
C54	←	ReliabMeans	0.880	.099	8.872	***	par_13
C61	←	StatBoards	1.000				
C50	←	StatBoards	0.970	.108	8.969	***	par_14
C48	←	StatSeats	1.000				
C46	←	StatSeats	1.161	.113	10.244	***	par_15
C14	←	Responsiveness	0.758	.078	9.712	***	par_26

Appendix 2. ClassroomQual Version 1 (for version 2, excluding C12) (Spanish)

F1	*Capacidad de reacción*
C12	Cuando en mis clases se requieren medios informáticos adicionales, los auxiliares los llevan con prontitud
C27	Los auxiliares logran resolver los problemas informáticos que surgen durante mis clases
C26	Los auxiliares resuelven todas las dudas que se tienen con respecto al uso de los medios informáticos
F2	*Distractores*
C57	La cantidad de estudiantes asignados al aula hacen que me sienta estrecho
C60	Al llegar a clase me toca desplazarme hacia otras aulas para buscar una silla
F3	*Iniciativa*
C1	El auxiliar se interesa por saber qué medios informáticos se requieren para las clases (computador, video beam, . . .)
C2	El auxiliar tiene la iniciativa de encender los medios informáticos (computador, video beam) requeridos en mis clases
C16	Los auxiliares se interesan por averiguar qué medios informáticos se necesitan en mis clases (computador, video beam)
F4	*Estado de medios tecnológicos*
C43	Cuando en mis clases se va a usar el video beam del aula, este no enciende
C44	Cuando en mis clases se va a usar el computador del aula, no enciende
F5	*Estado de sillas*
C46	Con frecuencia me toca usar sillas con asientos quebrados
C48	Con frecuencia me toca sentarme en sillas con asientos despegados
F6	*Estado de tableros*
C50	El estado de los tableros de las aulas no me deja ver con claridad la escritura
C61	El estado del tablero dificulta ver la tinta del marcador
F7	*Fiabilidad de medios tecnológicos*
C54	Con frecuencia el video beam se apaga inesperadamente en medio de mis clases
C55	Los medios informáticos de las aulas fallan con frecuencia

Appendix 3. ClassroomQual Version 1 (for version 2, excluding C12) (English)

F1	*Responsiveness*
C12	When my classes require additional computing resources, support personnel carry them promptly
C27	Support personnel can resolve the computer problems that occur during my classes
C26	Support personnel resolve all doubts concerning the use of the computer resources
F2	*Distractors*
C57	The number of students assigned to the class makes me feel cramped
C60	When I arrive in the classroom, I have to look into other classrooms to find a chair
F3	*Initiative*
C1	The support person is interested in knowing what computer resources are required for the classes (computer, projector, etc.)
C2	The support person takes the initiative to turn the equipment on (computer, projector, etc.), as required by my classes
C16	Support personnel are interested in finding out what computer resources are needed in my classes (computer, projector, etc.)
F4	*state of the technological means*
C43	When my classes are going to use the room's video beam, it does not turn on
C44	When my classes are going to use the room's computer, it does not turn on
F5	*State of the seats*
C46	I often have to use chairs with broken seats
C48	I often have to sit on chairs with loose seats
F6	*State of the boards*
C50	The state of the classroom boards prevent me from seeing the writing clearly
C61	The state of the board makes it difficult to see the marker ink
F7	*Reliability of the technological means*
C54	The projector often shuts down unexpectedly in the middle of my classes
C55	The classroom computer equipment fails frequently