CrossMark

Determinants of students' outcome: a full-fledged structural equation modelling approach

Mohammed Borhandden Musah[1] · Hairuddin Bin Mohd Ali[2] ·
Shafeeq Hussain Vazhathodi Al-Hudawi[1] · Lokman Mohd Tahir[1] ·
Khadijah Binti Daud[1] · Abdul Rahim Hamdan[1]

Abstract The vibrant demand for academic excellence in the twenty-first century has brought diverse determinants of students' outcome into play. However, few studies have validated the instruments and examined the mediating effect between exogenous and endogenous variables of the student outcome model. This study, therefore, investigates the psychometric properties of four scales: two modified subscales of SERVQUAL: quality of teaching and learning and quality of classroom, qualified instructional leader and student outcome. It also investigates student outcome determinants and the role of teaching and learning and classroom quality in mediating the relationship between qualified instructional leaders and student outcomes. The study adopts a full-fledged structural equation modelling approach and analyses data collected through a total of 450 questionnaires distributed among undergraduates of a public university in Malaysia. The findings demonstrate that student outcomes and qualified instructional leader instruments comprise three dimensions, whereas teaching and learning and classroom quality instruments comprise a single dimension each and support the composite reliability, convergence and discriminant validity of the scales. The results indicate that determinants of student outcome predict learning student outcomes. The results also demonstrate that teaching and learning quality strongly and positively mediates the relationship between qualified instructional leaders and student outcomes, whereas classroom quality positively but weakly mediates the relationship. The theoretical and practical implications of the study were also addressed.

Keywords Psychometric properties · Student outcome · Causal effect · Determinant · Structural equation modelling

Introduction

Academic excellence is key to the success of any educational institution (Ali and Musah 2012). The attributes "clear and focused mission, high expectation for success, instructional leadership, frequent monitoring of student progress, opportunity to learn and student time on task, a safe and orderly environment, and a conducive home school relationship", are partially responsible for improving academic excellence (Kunje et al. 2009: 4). Among these seven attributes, this study investigated the psychometric properties of instructional leadership in the context of qualified instructional leader (QIL) to determine how QIL predicts student outcome (SO) and the extent to which teaching and learning quality (QTL) and classroom quality (QC) mediate the relationship between QIL and SO.

The psychometric validation of a scale is a prerequisite for collecting information from participants (Gillham 2008; Pedhazur and Schmelkin 1991; Grimm and Yarnold 2006) because the relevancy, validity and reliability of the instrument are associated with accurate findings (Levine 2005; Said et al. 2011). Levine et al. (2006) argued that investigating and validating existing and previously assessed measures would provide valuable information and add to the generation of empirical knowledge. Furthermore, Messick

✉ Mohammed Borhandden Musah
mbmusaus@yahoo.com; borhandden@utm.my

[1] Department of Educational Foundation and Social Science, Faculty of Education, Universiti Teknologi Malaysia, Skudai Campus, 81310 Johor Bahru, Johor, Malaysia

[2] Department of Social Foundation and Educational Leadership, Kulliyyah of Education, International Islamic University Malaysia, Selangor, Malaysia

 Springer

(1989, p. 13) contended on the grounds of unitary theory that validity is an "integrated evaluative judgement of the degree to which empirical and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other models of assessment". Thus, establishing empirical and theoretical evidence of any instrument is critical prior to its usage in a given test.

However, due to the fact that higher education institutions (HEIs) have multiple outputs including responsibility for knowledge creation and dissemination, it has become more challenging for HEIs to bottom-line SO as a core business for their functionality (Marginson 2007; Coates 2010). The continuous demand for academic excellence through SO has been slating substantial pressure on HEIs to meet the needs of diverse stakeholders. Despite vibrant demand for academic excellence, recent literature discloses debates about whether teaching should be considered and promoted as a profession, or whether it should be deregulated and opened up to people without formal teacher preparation (Darling-Hammond and Youngs 2002). These debates highlight the importance of instructional leadership and encourage research in the teacher qualities directly linked to SO (Darling-Hammond and Youngs 2002).

With this background as a preliminary, this study assesses the construct validity and tests the psychometric properties of four subscales: QIL, QTL, QC and SO instruments. Second, the study investigates the determinants of the SO model: QIL, QTL and QC. Finally, it investigates the role of QTL and QC in mediating the relationship between QIL and SO.

Though factors of SO and SERVQUAL (except QIL) are widely used, there was neither evidence on their construct reliability nor construct validity (Mulford and Silins 2010, 2011). Therefore, this study tests the psychometric properties of QIL, QTL, QC and SO instruments. Thus, it is hypothesised that:

Hypothesis 1 Qualified instructional leader, student outcomes and SERVQUAL constructs are valid and reliable.

Determinants of student outcomes

Qualified instructional leader

Many studies have empirically endorsed the effects of QIL on classroom effectiveness and academic excellence (Cheng 1991 and 1994; Kunje et al. 2009; Lai et al. 2009). Lai et al. (2009) investigated high school students' achievement and found that the teacher quality significantly influences the learner's academic performance. In mathematics, Robinson (2009) found a moderate relationship ($r = 0.283$) between a teacher holding full mathematics certification and a student achievement in mathematics. Elsewhere, Cheng (1994) concluded that instructional leaders substantially influence the perceived physical quality of the classroom, social environment and student achievement in Hong Kong primary schools. Interestingly, it is assumed that QIL does not influence SO per se, but also other potential variables as (Cheng 1994; Abdullah 2005) implied. Thus, other than SO, QIL influences QC and QTL in the learning environments.

Cheng (1991, 1994) also argued that QC is hardly realised in isolation of the instructional leader's effectiveness. Oliver et al. (2011) reached the conclusion that the instructional leader's effectiveness improves QC. They further argued that instructional leader's effectiveness has significant and positive effects on minimising disruptive behaviours among students.

Educational theory suggests that the instructive quality is one of the most important determinants of QTL (Desimone et al. 2007; Mashburn et al. 2008). According to Akyeampong et al. (2012) and Lin et al. (2010), the quality of instructional leaders plays a significant role in attaining QTL. Empirical evidence showed that the effectiveness of teaching and learning is credited to the instructor's quality (Andrew and Schwab 1995; Bents and Bents 1990; Wang and Fwu 2007).

Quality of classroom

QC refers to the facilities for conducive teaching and learning. According to Owoeye and Yara (2011), QC among other school facilities strongly determines academic achievement. The availability, quality, adequacy and relevance of the classroom facilities influence efficiency and high academic outcomes (Oni 1992). Furthermore, Balgun (1982) argued that the promising outcome of education would be hardly realised in isolation of adequate teaching and learning facilities. A conducive environment equips students with potential skills in solving problems and develops a self-regulatory scientific attitude of learning. As a result, learners learn at their own pace (Owoeye and Yara 2011). A similar conclusion was reached by Earthman (2004) who argued that the quality of the physical environment of the classroom has significant effects on the student's academic achievement. According to Guo et al. (2010), classroom quality serves as a significant and positive predictor of the learner's achievements in preschool education.

Quality of teaching and learning

QTL is believed to have significant impacts on students' academic and social developments (Zammit et al. 2007). Furthermore, the findings of educational reform initiatives in several US states (North Carolina, Connecticut,

Kentucky, West Virginia, etc.) revealed that the quality of teaching accounts for the advancements on the academic achievement scale (Darling-Hammond 2000).

Interestingly, it is argued that QTL does not only influence the academic excellence of SO per se, but also mediates the relationship between QIL and SO. Similarly, it is also believed that QC does not influence SO per se, but goes further to mediate the relationship between QIL and SO. However, previous studies did not investigate the role of mediation effects on the endogenous variable, although such concerns were put forward by some researchers (Cheng 1994; Abdullah 2005). Cheng (1994), for instance, stressed that given the conceptual stand, the effects of the teacher's leadership on academic achievements may not be so direct. As such, one of the major contributions to this hypothesised model is its examination of mediation effects.

Furthermore, the findings of various studies disclosed diverse gender performance across SO determinants (Dayioglu and Turut-Asik 2007; Falch and Naper 2013; Wan Chik et al. 2012; Wang and Staver 1997). Thus, males and females perform differently on SO scales. Wan Chik et al. (2012) investigated undergraduate nursing students and found gender to vary significantly across academic performance. Similarly, Dayioglu and Turut-Asik (2007), Falch and Naper (2013), Severiens and ten Dam (2011) and Sheard (2009) concluded that female students outperformed their male counterparts on achievement scales in tertiary institutions (Fig. 1).

This study tests the effects of QC and QTL as mediators, as well as assessing the direct effects of QIL, QTL and QC on SO in the context of tertiary education. It is, therefore, hypothesised that:

Hypothesis 2 Qualified instructional leader directly determines student outcomes.

Hypothesis 3 Quality of teaching and learning determines student outcomes.

Hypothesis 4 Quality of classroom determines student outcomes.

Hypothesis 5 Qualified instructional leader determines quality of teaching and learning.

Hypothesis 6 Qualified instructional leader determines quality of classroom.

Hypothesis 7 Quality of teaching and learning positively mediates the relationship between qualified instructional leader and student outcomes.

Hypothesis 8 Quality of classroom positively mediates the relationship between qualified instructional leader and student outcomes.

Hypothesis 9 Student outcomes model is gender invariant.
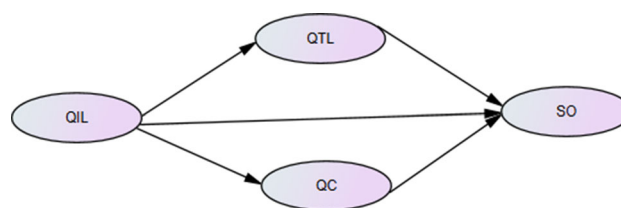
Fig. 1 Student outcome model, with QTL and QC mediating the relationships between QIL and SO. *Source* Mulford and Silins 2011; Parasuraman et al. 1985; Robinson et al. 2008; Zammit et al. 2007

Method

Sample

A total of 450 undergraduates at one tertiary institution participated in the present study. The participants were full-time students undergoing tertiary education in the fields of engineering, economic, Islamic studies, information and communication technology, law, education and architecture. The researchers personally distributed surveys to randomly sampled undergraduate students of the sampled public university in Kuala Lumpur. Only final-year undergraduates in eight faculties of the sampled public university were selected. Of the 300 returned questionnaires, 19 were discarded because they were incomplete resulting in a total of 281 questionnaires retained and analysed. This indicated a response rate of 62.44 %.

Instrument

The QIL instrument used in the study was developed by the researchers based on the literature on instructional leadership and related studies (Cheng 1991, 1994; Lai et al. 2009; Robinson et al. 2008) because appropriate items or standardised instruments relevant to measure the construct of the QIL were unavailable.

The initial pool of 11 items generated was content and construct validated using various methods such as subject matter experts (SMEs) judgement and the use of Cronbach's alpha to check the internal consistency of items. Items with a low Cronbach's alpha (<.70) were discarded from further analysis because they were not entirely reliable (Litwin 1995). Furthermore, assessment of composite reliability index (CRI) and average variance extracted (EVA) were also calculated.

QC and QTL instruments were adapted from a modified SERVQUAL dimensions (Parasuraman et al. 1985; Stodnick and Rogers 2008). Given the context of this study, four dimensions (tangibility, responsiveness, reliability and assurance) were selected.

Despite the fact that SERVQUAL instruments were widely used since their development, a thorough empirical

test of their psychometric properties has not been carried out other than in a study by Abdullah (2005), who employed the SEM approach to evaluate the constructs. Abdullah (2005) study did not provide rigorous evidence of construct validity and reliability in terms of AVE and CRI evidence.

With regard to the SO construct, accountability and evaluation, student empowerment and social skill development were adapted from successful principalship inventory. This instrument has been widely used in the successful school context. It is worth noting, however, that other than reliability estimates, fixed effects, deviance and variance of components, studies of Mulford and Silins (2010, 2011) did not reveal evidence of psychometric properties. Given the above arguments, this study tested the psychometric properties of QIL, QTL, QC and SO instruments. Responses to all items were made on a seven-point Likert scale anchored with 1 = very strongly agree to 7 = very strongly disagree.

Content validity

Since the QIL instrument was developed by the researchers, content validity was conducted prior to data collection to ensure that a detailed description of the content domain is captured. To perform this assessment, ten survey questionnaires were administered to SMEs in the area of teaching and learning in tertiary education setting. The participants were purposively selected based on their expertise in the subject area and asked to assess the instrument in terms of its relevancy and representativeness of the content domain. The selection of this particular group for this special task was based on their close engagement in monitoring and improving instructional quality. The ten participants concluded that the 11 items constituting QIL are clear and had captured its core elements. However, obtaining a sound instrument per se is not enough to generate reliable information from the participants; it should be equally assured that the instrument is valid and reliable given its construct. As such, rigorous analyses were further performed to assess the construct validity and reliability of the instrument.

Results and discussion

The results of the descriptive frequencies indicated that 154 (54.8 %) of the respondents were males, while 127 (45.2 %) were females. Internal consistency of the instruments was then assessed revealing an overall Cronbach's alpha for QIL .71, QC .87, QTL .92 and SO .90, respectively. This provides evidence that the items were internally consistent and reliable.

Table 1 Extreme values of Mahalanobis D^2

Mahalanobis distance	Case number		Value
Highest	1	156	18.13003
	2	183	16.42000
	3	219	14.13078
	4	157	12.67833
	5	21	12.59523
Lowest	1	11	.05862
	2	6	.11929
	3	109	.15079
	4	46	.15079
	5	228	.15607

The dataset was then examined for univariate and multivariate outliers. The results disclose that none of the cases (though two cases 145, ± 2.35 and 156, ± 2.22 were relatively high, but less than the suggested ratio) exhibits a Z-score greater than ±2.5, which indicates lack of univariate extreme case in the data (Hair et al. 2010; Meyers et al. 2006).

Multivariate outliers were then inspected by computing Mahalanobis distance for each case on the four variables, of which none was detected ($p > .001$). Because none of the critical values associated with any of the cases ≥ 18.467, based on the χ^2 criterion, we concluded that no multivariate outliers are detected in association with the four variables included in the analysis. Table 1 presents the details.

Construct validity

Since we sought to identify the psychometric properties of the constructs, a more rigorous SEM-based approach of confirmatory factor analysis (CFA) was used to validate the constructs understudy. CFA would be the best choice to validate instruments that have been fully developed and whose factor structures have been validated (Byrne 2010; Levine et al. 2006). SERVQUAL and successful school principalship inventories, as widely used, met this criterion.

Validating measurement models

Two CFA models were assessed, out of which one was used for the QIL, QC and QTL factors, and the other was used for the SO factors. The first CFA model revealed a good fit to the data. The coefficients of all items measuring this construct were high, and all were above the critical ratio cut score of 1.96. This is an indication of the practical significance of indicators (Hair et al. 2010). The results of fit indices showed that the hypothesised model received a good fit to the sample data: $\chi^2 = 230.721$, $df = 116$,

comparative fit index (CFI) = .95, Tucker Lewis index (TLI) = .94, incremental fit index (IFI) = .91 and root-mean-square error of approximation (RMSEA) = .06. Interestingly, the squared multiple correlations (SMCs), which indicate how well the observed variables serve as measures of the latent variables, were also investigated. It is worth noting that the SMCs' values of the QTL, QC and QIL measurement models had fulfilled the requirement ($\geq$.25). The values ranged from .28 through to .77. This provides reasonable evidence of the reliability of the parameters' estimates.

In addition, the model indicates that the parameters were free from issues of offending estimates. The estimates ranged from .53 through to .88, indicating psychometric evidence of instrument quality, thereby constituting a satisfactory level of data analysed at the item level. Moreover, the composite inter-factor correlations yielded (.52, .52 and .66) relation among three factors, which is an indication of discriminant validity (Kline 2010). Thus, none of the inter-factor correlations reached or exceeded .85. Table 2 presents the details.

A second CFA analysis was also performed to evaluate the SO construct, which contains three factors. The results of the CFA model revealed a good fit to the data. Fit indices demonstrate that the hypothesised measurement model received a good fit to the data: $\chi^2 = 228.626$, $df = 116$, CFI = .95, TLI = .94, IFI = .93 and RMSEA = .06. However, loadings of some items (SE3 .41 and SE4 .39) fell below $\geq$.50, indicating a need to revise the hypothesised measurement model of SO construct.

Revised model

Modification indices (MIs) need to be inspected for possible reasons leading to low loading of the two items failed to meet inclusion criteria. The inspection yielded a relative

multicollinearity between Items SE3 and SE4. In other words, the two items were highly correlated, which indicates the possibility of similar content holds by them. Consequently, Item SE4 was dropped. Although Item SE3 fell <.50, it was retained for theoretical and practical reasons. Moreover, it is worth noting that the values of SMCs of the SO measurement model had fulfilled the requirement ($\geq$.25) except for Item SE3 (.18). The values ranged from .18 through to .79. This provides reasonable evidence of the reliability of the loadings. Table 3 demonstrates the details.

Construct validity and reliability

Given instrument validation as one of the main purposes of this study, subscales of SO, QIL and modified version of SERVQUAL were further evaluated. CRI, AVE and discriminant validity for each construct were further assessed. According to Fornell and Larcker (1981), evidence of convergent validity is obtained if the AVE is at least $\geq$.50. Furthermore, evidence of construct reliability is established if CRI of each factor is $\geq$.70.

Interestingly, the results demonstrate that the measures are reasonably converged on their respective factors. Meanwhile, all factors exceeded the recommended threshold CRI of $\geq$.70 and AVE of $\geq$.50, indicating the attainment of construct validity on the scale of adequacy and appropriateness (Messick 1989). These results supported *Hypothesis 1* with the finding that constructs of qualified instructional leader, quality of teaching and learning, quality of classroom and student outcome hold evidence of construct validity and reliability. On a similar note, the results allow us to assess the extent to which measures of one construct are empirically distinct from each other (Bagozzi and Burnkrant 1985). That is to say, discriminant validity should be performed.

The results show substantial evidence of discriminant validity pertaining to all factors (Byrne 2010; Fornell and Larcker 1981). Although the squared inter-factor correlation of SE factor seems to lack discriminant evidence, it had fulfilled the criteria from different perspectives. According to Kline (2010), if inter-correlations of a set of variables that are presumed to measure different factors are not too high ($\leq$.85), evidence of discriminant validity is established. However, if the opposite is true $\geq$.90, then evidence of discriminant validity cannot be claimed (Kline 2010). In a nutshell, the measurement models demonstrate adequate reliability, convergence and discriminant validity. Table 4 demonstrates the details.

Adequacy of the causal structure of student outcomes model

The SEM results summarise the causal effects of the SO model. The confirmatory modelling yielded consistency of

Table 2 QTL, QC and QIL factor loadings and goodness-of-fit criteria for the sample data

Item	QTL			QC	QIL
	AS	RES	REL	CR	QL
1	.85	.82	.82	.82	.53
2	.72	.79	.64	.88	.59
3	.72	.76	.64	.80	.77
4	.73				.69

χ^2	df	IFI	TLI	CFI	RMSEA	90 % CI
Goodness-of-fit criteria						
230.721	116	.91	.94	.95	.06	.05–.08

AS assurance, *RES* responsibility, *REL* reliability, *CR* classroom, *QL* qualified leader

Table 3 SO factor loadings and goodness-of-fit criteria for the sample data

Item	SSD (SO1)	AE (SO2)	SE (SO3)
Student outcome			
1	.85	.82	.77
2	.80	.85	.75
3	.83	.89	.42
4	.79	.71	
5	.78	.72	
6	.70		
7	.78		
8	.72		

χ^2	df	IFI	TLI	CFI	RMSEA	90 % CI
Goodness-of-fit criteria						
211.695	101	.96	.94	.95	.06	.05–.07

Table 4 Construct reliability and validity of the measurement models

Latent construct	AVE	Squared inter-factor r	CRI
QIL	.56	.27	.66
QC	.64	.27	.87
QTL	.56	.43	.92
SE (SO3)	.82	.57	.81
AE (SO2)	.53	.33	.89
SSD (SO1)	.61	.36	.92

Composite reliability index (CRI) formula $= \left(\sum foctor\ loadingd \right)^2 / \left(\sum foctor\ loadingd \right)^2 + \sum \varepsilon j$. Average variance extracted (AVE) formula $= \left(\sum foctor\ loadingd \right)^2 / \left(\sum 1 - foctor\ loadingd \right)^2 + \sum \varepsilon j$.

the hypothesised causal relationships with the data: χ^2 (487) = 874.211, $p = .001$, CFI = .92, TLT = .92, RMSEA = .05 and $\chi^2/df = 1.795$. All these fit indices satisfied their critical cut scores. The results, therefore, indicated a good fit of the SO model to the data.

Moreover, the level of discrepancy between the hypothesised model and the data, divided by the degrees of freedom, demonstrated substantive fit at χ^2/df 1.795. Given the guideline of the statisticians (Byrne 2010; Hair et al. 2010), the complexity of this model, CFI threshold of more than .90 and RMSEA threshold of .07 reflect a good fitting model. The values of CFI (.92) and TLI (.92) suggest more evidence to support the goodness of the hypothesised model fit. Likewise, the RMSEA, with its CI of the LO and HI, also fell within the desired zone (LO .05 HI .06), providing additional evidence of model acceptance (Chen et al. 2008). In addition, the model is consistent with the valid measurement, since the absence of model

contamination-related issues confounded with error term connection is established (Levine 2005). Figure 2 depicts the details.

The parameter estimates of the SO hypothesised model were free from offending estimates. All path coefficients of the causal structure were statistically significant at .005 levels and were of practical importance (except QC → SO), since the smallest value of the standardised path coefficient was .21. Moreover, the SMCs were also investigated. It is worth noting that the values of SMCs of the SO hypothesised model had fulfilled the requirement ($\geq.25$) for all indicators other than SE3 (.18). The values ranged from .18 (SE3) through to .78 (AE2). This provided substantive evidence to explain the variance in the 33 indicators of the SO model.

Standardised causal effects of students' outcome model

The study used SEM with a significant level of .005 to test the directional effects of the hypotheses 2, 3, 4, 5, 6, 7, 8 and 9. Using the standardised causal effects, the direct, indirect and total effects of the three latent constructs on SO were examined.

According to Fig. 2, the total standardised effect of QIL → QTL was .72 and it was statistically significant, indicating that for each unit increase in QIL, there will be a .36 unit increase in the QTL, controlling other variables in the model. This finding agrees with those of Bents and Bents (1990), Andrew and Schwab (1995), Wang and Fwu (2007), Desimone et al. (2007), Mashburn et al. (2008), Lin et al. (2010) and Akyeampong et al. (2012) that teacher professional qualities strongly influence QTL. This result supported *hypothesis 5,* with the finding that qualified instructional leader determines quality of teaching and learning.

Figure 2 also shows that the total standardised effect of QIL → QC was .60 and was statistically significant, indicating that for each unit increase in QIL, there will be a .30 unit increase in the QC, controlling other variables in the model. This finding corresponds with those of Oliver et al. (2011) and Cheng (1991, 1994) who concluded that qualified individual instructor substantially determines the perceived physical aspect of QC. Furthermore, the results supported *hypothesis 6.* This is an indication that qualified instructional leader influenced the quality of classroom in the tertiary education context.

Furthermore, the results show that the standardised total effect of QIL → SO was .91 and was statistically significant, indicating that for each unit increase in QIL, there will be a .46 unit increase in the SO, controlling other variables in the model. This strong direct causal effect indicates that QIL strongly influences SO. This finding is
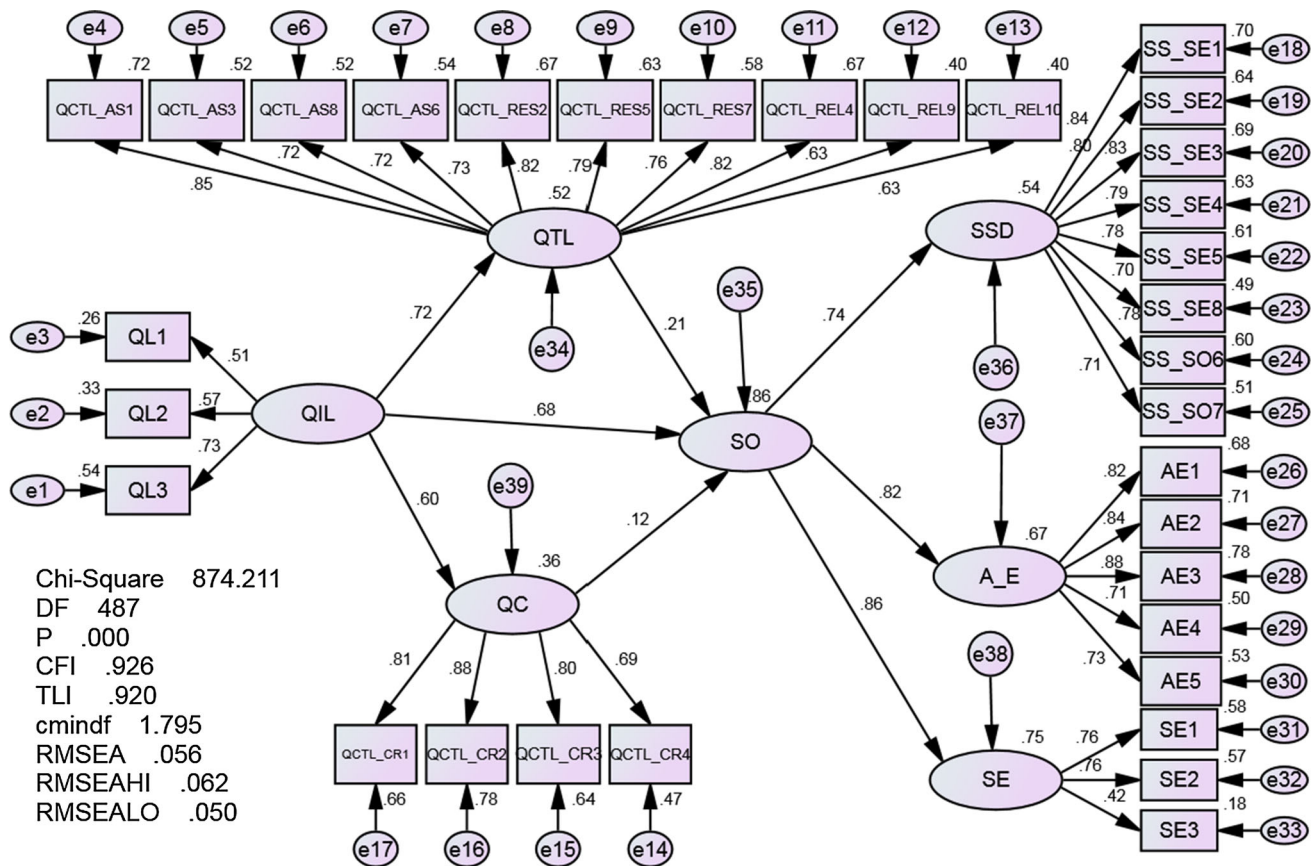
Fig. 2 Generated fit indices of the hypothesised model of student outcome

consistent with Lai et al. (2009) who reached the conclusion that QIL significantly influences the learner's academic performance. In addition, this result showed that *hypothesis 2,* with the finding that qualified instructional leader directly determines student outcomes, is supported.

As shown in Fig. 2, the direct path coefficient between QTL → SO was moderate (.21) and was statistically significant, indicating that for each unit increase in QTL, there will be a .12 unit increase in SO, controlling other variables in the model. This finding is consistent with those of Darling-Hammond (2000) and Zammit et al. (2007) that pedagogical quality and conducive environment strongly determine learner's academic outcomes. Consequently, this result supports *hypothesis 3,* which states that quality of teaching and learning determines student outcomes.

Moreover, the direct path coefficient between QC → SO was statistically insignificant (.12), indicating that for each unit increase in QC, there will be a .6 unit increase in SO, controlling other variables in the model. This finding contradicted with the findings of Balgun (1982), Oni (1992), Earthman (2004), and Owoeye and Yara (2011) where it was discovered that QC significantly influences SO. The result also indicated that the

determinant of SO (i.e., QC) had a weak predictive causal effect which accounted for the direct effect of QC on the learning outcome of undergraduates at the sampled institution. Consequently, this result did not support *hypothesis 4* that the quality of classroom directly determines SO.

The indirect effects of QIL on SO through both QTL and QC were also investigated. The estimation method for practical importance states that if the standardised direct effect of X_1 on Y_2 (i.e., QIL × QTL) is ≥.08, then the significance of the indirect effect is held and vice versa (Kline 2010). Given this convention, Sobel's (1982) test method was used to assess statistical significance. In other words, indirect effects of QIL on SO through the mediating variables in which path coefficient for the relationship between QIL and mediators and standard errors of the relationships of both are calculated based on the significance of *p* value.

Given the formulas suggested by Kline (2010) and Sobel (1982) in calculating indirect effects, the magnitude of the indirect effect of QIL on SO through QTL was assessed. The calculation revealed that the result (.72 × .21 = .15) was far greater than .08. This result indicates that qualified instructional leader practically and partially determines

student outcomes indirectly through quality of teaching and learning.

In examining the statistically significant mediation effect using Sobel's (1982) test method, several steps were taken to assess QTL as a mediator of the relationship between QIL and SO. First, QIL was regressed onto QTL ($\beta = .72$, SE = .10). Secondly, QTL was regressed onto SO ($\beta = .21$, SE = .06). All four statistical values were entered into an online version 3.0 of Sobel's tests (Soper 2009) to determine the statistical significance of the mediating variable. The analysis yielded a Sobel's test statistics of 3.1477, $p < .001$, which indicated that QTL exhibits a statistically significant mediation effect. Furthermore, these results addressed and supported *hypothesis 7* that quality of teaching and learning mediates the positive relationship between qualified instructional leader and student outcomes in the sampled institution.

In addition, the indirect causal effect calculation for QIL on SO, through QC, revealed a result of .07 (.60 × .12). This magnitude is below ≥.08, indicating insignificant practical indirect causal effects between QIL and SO in the studied institutional context. QIL was then regressed onto QC ($\beta = .60$, SE = .10), and QC was regressed onto SO ($\beta = .12$, SE = .06) to evaluate statistical significance. The analysis yielded a Sobel's test statistic of 1.9873, $p > .057$, which indicated that QC exhibits an insignificant statistical mediating effect. These results addressed *hypothesis 8,* which states that quality of classroom mediates a positive relationship between qualified instructional leader and student outcomes. The mediation was observed to be weak and insignificant. This finding indicates that qualified instructional leader weakly determines the undergraduate's learning outcome through quality of classroom in the context of the sampled institution. Interestingly, the analysis revealed that the three exogenous variables collectively explained 86 % of the variability of the SO. The study provided evidence of the presence of a significant causal relationship among the variables investigated. Table 5 depicts the results of standardised causal effects of the SO model.

Gender invariance of student outcome model

The study examines the structural invariance of the SO model across male and female groups. To test gender invariance, a simultaneous analysis of male ($n1 = 155$) and female ($n2 = 126$) samples was conducted using the following steps. First, without constraining the structural paths, the results derived a baseline Chi-square value. Next, the structural paths (QIL → QTL; SO and QIL → QC; SO) were constrained to be equal for male and female groups. The analysis of this constrained SO model produced another χ^2 value which was then tested against the baseline value for statistically significant differences.

Table 5 Summary of the standardised causal effects of student outcome

Outcome	Determinant	Direct	Indirect	Total
Causal effects				
QTL	QIL	.72	.22	.72
QC	QIL	.60	.01	.60
SO	QIL	.68		.91
SO	QTL	.21	–	.21
SO	QC	.12	–	.12

Table 6 Results of the moderating effect

Gender	χ^2	Df	Critical value	χ^2 Change
Unconstrained	1743.264	976	9.49	4.28
Constrained	1747.545	980		

The invariance test across male and female groups resulted in a statistically insignificant change in the χ^2 value, Chi-square ($df = 4$) = 4.28, $p \geq .005$. This simply means that the difference in the χ^2 values between the unconstrained and the constrained models did not produce a poorer fit model. The path coefficients did not vary significantly across gender. Therefore, it can be concluded that gender did not interact with the exogenous variables to influence students' academic learning outcome; hence, gender is not a moderating variable. This result addressed and rejects *hypothesis 9* and states that gender did not moderate the SO model of the undergraduates of the investigated tertiary institution. This finding was inconsistent with the findings of previous studies, though dichotomous; Dayioglu and Turut-Asik (2007), Falch and Naper (2013), Wan Chik et al. (2012), Wang and Staver (1997) in favour of males and Dayioglu and Turut-Asik (2007), Falch and Naper (2013), Severiens and ten Dam (2011) and Sheard (2009) in favour of females found variances of gender across SO determinants. The insignificant gender interaction found in this study could be explained that qualified instructional leader's gender does not potentially contribute to undergraduate student learning outcomes. Thus, the expertise of qualified instructional leader rather than gender potentially contributes to learning outcomes. Table 6 depicts the details.

Conclusion

The study found a strong direct causal effect of qualified instructional leader on student outcomes in the context of the sampled public university. We conclude that the exogenous variable included in the model strongly

predicted students' learning outcome. The findings established that qualified instructional leader strongly determined undergraduate learning outcome through quality of teaching and learning, but weakly determined learning outcome through quality of classroom in the context of the public university sampled. This indicates that the issue of poor academic performance of undergraduates can be overcome through employing standard and updated teaching methods that facilitate the learning process. The above weak predictive causal effect of quality of classroom on the learning outcome of undergraduates in the sampled public university projects the conclusion that undergraduates are not really particular about quality of classroom in relation to their performance on the academic scale.

Furthermore, the study supports that gender did not interact with the exogenous variables to influence students' outcomes; hence, gender is not a moderating variable. In addition, the results indicated that quality of teaching and learning exhibited the second largest direct causal effect compared to the other exogenous variables included in the study.

Taken together, the results demonstrated that both self-constructed and modified selected subscales have sound psychometric properties and valid, reliable factor structures and therefore contributed to the literature of instructional leadership. Although further research is necessary to replicate the present findings and provide additional evidence of the psychometric properties, the construct, convergent and discriminant validities were established in this study. Undoubtedly, the sample size was relatively small when compared to the student population in the sampled institution. Similarly, the study was also limited to only one university. Future studies should test these results using larger sample sizes, and survey many HEIs.

The findings have implications for practice and pedagogy. Since the effectiveness of qualified instructional leader is relative given the mediation role, a university management may consider introducing intervention programmes or professional training to update and increase instructional quality and effectiveness. In addition, institutional support from the university management in terms of incentives and facilities is a crucial element that will influence the instructor's effectiveness.

The findings also revealed adequacy of the self-developed qualified instructional leader instrument, which can be used as a means to predict the impact of the instructional leader's quality on ameliorating student outcomes.

Limitations

Despite the fact that the findings initiate preliminary, valid and reliable empirical findings, there were some limitations. First, the study sampled only one public university in Kuala Lumpur, excluding 68 public and private HEIs nationwide. Second, the study exclusively investigated final-year undergraduates' perceptions with regard to determinants of student outcomes. There was no inclusion for lecturers or management perceptions in the study. Thus, the results should be interpreted with caution. This suggests that future research in Malaysia should diversify participants (teaching staff and management) and include more HEIs. Finally, the use of a quantitative approach in data collection and analysis might be another limitation to the findings. Thus, future studies should use a mixed-method approach to study the variables for more robust conclusions.

References

Abdullah, F. (2005). Measuring service quality in higher education: HEdPERF versus SERFPERF. *Marketing Intelligence and Planning, 24*(1), 31–47.

Akyeampong, K., Lussier, K., Preyor, J., & Wetbrook, J. (2012). Improving teaching and learning basic math and reading in Africa: Does teacher preparation counts? *International Journal of Educational Development, 33*(3), 272–282.

Ali, H. M., & Musah, M. B. (2012). Investigation of Malaysian higher education quality culture and workforce performance. *Quality Assurance in Education, 20*(3), 289–309. doi:10.1108/09684881211240330.

Andrew, M., & Schwab, R. L. (1995). Has reform in teacher education influenced teacher performance? An outcome assessment of graduates of eleven teacher education programs. *Action in Teacher Education, 17*, 43–53.

Bagozzi, R. P., & Burnkrant, R. E. (1985). Attitude organisation and the attitude behaviour relation: A reply to Dilon and Kumar. *Journal of Personality and Social Psychology, 49*, 47–57.

Balgun, T. A. (1982). Improvisation of science teaching equipment. *Journal of the Science Teachers Association, 20*(2), 72–76.

Bents, M., & Bents, R. (1990). Perceptions of good teaching among novice, advanced beginner and expert teachers. In *Paper presented at the annual meeting of the American Educational Research Association*, Boston, MA, USA.

Byrne, B. M. (2010). *Structural equation modelling with AMOS: Basic concepts, applications and programming* (2nd ed.). USA: Taylor and Francis Group LLC.

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cut-off points in RMSEA tests statistic in structural equation models. *Sociological Methods and Research, 36*(4), 462–494. doi:10.1177/0049124108314720.

Cheng, Y. C. (1991). *Function and effectiveness of education*. Hong Kong: Wide Angle Press.

Cheng, Y. C. (1994). Teacher leadership style: A classroom-level study. *Journal of Educational Administration, 32*(3), 52–71.

Coates, H. (2010). Defining and monitoring academic standards in Australian higher education. *Higher Education Management and Policy, 22*(1), 2–17.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives, 8*(1), 1–44.

Darling-Hammond, L., & Youngs, P. (2002). Defining "highly qualified teachers": What does "scientifically-based research" actually tell us? *Educational Researcher, 31*(9), 13–25.

Dayioglu, M., & Turut-Asik, S. (2007). Gender differences in academic performance in a large public university in Turkey. *Higher Education, 53*(2), 255–277. doi:10.1007/s10734-005-2464-6.

Desimone, L. M., Smith, T. M., & Frisvold, D. (2007). Has NCLB improved teacher and teaching quality for disadvantaged students? In A. Gamoran (Ed.), *Standards-based reform and the poverty gap: Lessons for 'no child left behind'*. Washington, DC: Brookings Institute.

Earthman, G. I. (2004). Prioritization of 31 criteria for school building adequacy. In *American Civil Liberties Union Foundation of Maryland*. Retrieved on November 20, 2012. http://www.aclu-md.org/aTop%20Issues/Education%20Reform/EarthmanFinal10504.pdf.

Lai, F. Sadoulet, E. & Janvry, A. (2009). The contributions of school quality and teacher qualifications to student performance: Evidence from a natural experiment in Beijing middle schools. Retrieved on 14 May, 2014 from https://gspp.berkeley.edu/assets/uploads/research/pdf/SchoolEffectJHR.pdf.

Falch, T., & Naper, L. R. (2013). Educational Evaluation Schemes and gender gaps in student achievement. *Economics of Education Review, 36*, 12–25.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50.

Gillham, B. (2008). *Developing a questionnaire* (2nd ed.). London: Continuum International Publishing Groups.

Grimm, L., & Yarnold, P. R. (2006). *Reading and understanding more multivariate statistics*. Washington DC: American Psychological Association.

Guo, Y., Piasta, S. B., Justice, L. M., & Kaderavek, J. N. (2010). Relations among preschool teachers' self-efficacy, classroom quality and children's language and literacy gains. *Teaching and Teachers Education, 66*, 1095–1103.

Hair, F., Anderson, E., Tatham, L., & Black, C. (2010). *Multivariate data analysis* (7th ed.). New Jersey: Pearson Prentice Hall.

Kline, R. B. (2010). *Principles and practices of structural equation modelling* (3rd ed.). New York: The Guilford Publications Inc.

Kunje, D., Selemani-Meke, E., & Ogawa, K. (2009). An investigation of the relationship between school and pupil characteristics and achievement at the basic education level in Malawi. *CICE Hiroshima University, Journal of International Cooperation in Education, 12*(1), 33–49.

Levine, T. R. (2005). Confirmatory factor analysis and scale validation in communication research. *Communication Research Reports, 2*(4), 331–338.

Levine, T. R., Hullett, C. R., Turner, M. M., & Lapinski, M. K. (2006). The desirability of using confirmatory factor analysis on established scales. *Communication Research Reports, 23*(4), 309–314.

Lin, R., Xie, J., Jeng, Y. C., & Huang, S. (2010). The relationship between teacher quality and teaching effectiveness perceived by students from industrial vocational high schools. *Asian Journal of Arts and Sciences, 1*(2), 167–187.

Litwin, M. S. (1995). *How to measure survey reliability and validity*. London: Sage Publications Inc.

Marginson, S. (2007). University mission and identity for a postgraduate public era. *Higher Education Research and Development, 26*(1), 131–142. doi:10.1080/07294360601166851.

Mashburn, A. J., Pianta, R. C., Barbarin, O. A., Bryant, D., Hamre, B. K., & Downer, J. T. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79*, 732–749.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–100). Washington DC: American Council on Education.

Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied multivariate research: Design and interpretation*. London: SAGE Publications Inc.

Mulford, B., & Silins, H. (2010). Organizational learning in schools. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International Encyclopedia of Education* (3rd ed., pp. 143–150). Oxford: Elsevier.

Mulford, B., & Silins, H. (2011). Revised models and conceptualisation of successful school principalship for improved student outcomes. *International Journal of Educational Management, 25*(1), 61–82.

Oliver, R. M., Wehby, J. H., & Reschly, D. J. (2011). Teacher classroom management practices: Effects on disruptive or aggressive student behaviour. *Campbell Systematic Reviews, 2011*, 4. doi:10.4073/csr.2011.4.

Oni, J. O. (1992). Resource and resource utilisation as correlates of school academic performance, Unpublished Ph.D. Thesis, University of Ibadan.

Owoeye, J. S., & Yara, P. O. (2011). School facilities and academic achievement of secondary schools of agricultural science in Ekiti, Nigeria. *Asian Social Science, 7*(7), 64–74.

Parasuraman, A., Zeithaml, V., & Berry, L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing, 49*(4), 41–50.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design and analysis: An integrated approach*. Hillsdale, NJ: Lawrence Erlbaum Associates Publisher.

Robinson, V. M. J., Llyod, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of differential effects of leadership types. *Educational Administration Quarterly, 44*(5), 635–674.

Robinson, D. I. (2009). Teacher quality as a factor of student achievement: How does the type of teacher certification correlate with student mathematics achievement? In *Phi delta kappa international, 2009 summit on quality educator recruitment and retention*.

Said, H., Badru, B. B., & Shahid, M. (2011). Confirmatory factor analysis (CFA) for testing validity and reliability of instrument in the study of education. *Australian Journal of Basic and Applied Sciences, 5*(12), 1098–1103. Retrieved on March 27, 2014. From http://goo.gl/GM5rWu.

Severiens, S., & ten Dam, G. (2011). Leaving college: A gender comparison in male and female-dominated programs. *Research in Higher Education,*. doi:10.1007/s11162-011-9237-0.

Sheard, M. (2009). Hardiness commitment, gender, and age differentiate university academic performance. *The British Journal of Educational Psychology, 79*(1), 189–204.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology 1982* (pp. 290–312). Washington, DC: American Sociological Association.

Soper, D. (2009). Sobel test calculator for the significance of mediation (Software). Retrieved November 20, 2013, from http://www.danielsoper.com/statcalc.

Stodnick, M., & Rogers, P. (2008). Using SERVQUAL to measure the quality of the classroom experience. *Decision Science Journal of Innovative Education, 6*(1), 113–115.

Wan Chik, W. Z., Salamonson, Y., Everett, B., Ramjan, L. M., Attwood, N., Weaver, R., et al. (2012). Gender difference in academic performance of nursing students in a Malaysian university college. *International Nursing Review, 59*, 387–393.

Wang, H. H., & Fwu, B. J. (2007). In pursuit of teacher quality in diversity: A study of the selection mechanisms of new secondary teacher education programmes in Taiwan. *International Journal of Educational Development, 27*(2), 166–181.

Wang, J., & Staver, J. R. (1997). An empirical study of gender differences in Chinese students' science achievement. *The Journal of Educational Research, 90*(4), 252–255. doi:10.1080/00220671.1997.10544579.

Zammit, K., Sinclair, C., Cole, B., Singh, M., Costley, D., a'Court, L., & Rushton, K. (2007). *Teaching and learning for quality Australian schools: A review and synthesis of research-based knowledge*. Australia: Australian Institute for Teaching and School Leadership LTD.