Online Peer Evaluation for Assessing Perceived Academic Engagement in Higher Education

Semiral Oncu
Uludag University, TURKEY

Many institutions monitor academic engagement to investigate student achievement and institutional performance. Relying only on self-reports is prone to misjudgment. Peer evaluation through teamwork has the potential to substitute for measuring engagement, which has not been emphasized in the literature. This study examines whether peer evaluation can represent perceived academic engagement with coursework, operationalized as "active learning" and "paying attention." Data were collected through an online peer evaluation system. Participants were 215 undergraduate and graduate students from three universities. Structural equation modeling (SEM) was used to test hypotheses. Results show that peer evaluation is statistically predictive of perceived academic engagement. Peers give high scores to their counterparts conforming to the literature. Students are more engaged than overall benchmarks. Results suggest that creating more opportunities for teamwork in the educational environment and efficient ways to incorporate peer evaluation through means such as social networks and online learning management systems are viable venues for improvement.

Keywords: Perceived academic engagement, active learning, collaboration, paying attention, peer evaluation.

INTRODUCTION

Literature shows that there is a positive link between academic engagement and achievement, although moderate in size (Carini, Kuh, & Klein, 2006; Johnson & Butts, 1983). Engagement is also linked to academic well-being (Newmann, 1992). It is directly associated with constructivist learning principles in that learning is constructed by and is special to the individual. Participation in educationally purposeful activities directly affects the quality of learning. Although in the constructivist perspective, learning is argued as well to be facilitated socially and influenced greatly by the context, the bulk is on the shoulders of the individual learner (Coates, 2005, p. 26) and how much self-directed

effort (Macdonald, 2004) is put in the learning activities. Therefore, academic engagement is considered to be a central factor in educational settings. However, the fact that students are not as engaged as they should be is a growing concern (Marks, 2000; National Survey of Student Engagement [NSSE], 2006).

Engagement is a process and an outcome; it is multidimensional and dynamic in nature (Annetta, Minogue, Holmes, & Cheng, 2009). It is, hence, assessed during or after a target program, semester, or an academic year through questions that are after-the-fact. Of the several different techniques measuring engagement, most rely on self-reported surveys, where students rate their *own* performance. While this has been reportedly employed in educational policy decisions (for example AUSSE Australasian Survey of Student Engagement in Australia and New Zealand, NSSE in the US and Canada, and SASSE South African Survey of Student Engagement), there is a need for conceptualizing engagement with multiple indicators and with a theory that takes into account multi dimensions or views (Krause, 2005). There is need to find better ways to assess students, especially online

Correspondence to: Semiral Oncu,
Department of Computers and Instructional Technology Education, Faculty of Education,
Uludag University, 16059 Bursa, TURKEY
E-mail: semiral@uludag.edu.tr

State of the literature

- Academic engagement is linked to academic well-being, but most engagement instruments rely only on student self-reports, which are prone to misjudgment.
- Peer evaluation, relying on outsiders' view can potentially help measure student engagement.
- There is no study, if not few, that investigates the relationship between perceived academic engagement and peer evaluation. This study highlights the association between the two performance indicators using an online peer evaluation tool.

Contribution of this paper to the literature

- Peer evaluation scores successfully predict perceived academic engagement, operationalized as active learning and paying attention.
- Peers give high scores to their counterparts conforming to the literature.
- Students are more engaged than overall benchmarks found in the literature.

(Oncu & Cakir, 2011). Relying only on self-reports of students is prone to misjudgment (Podsakoff & Organ, 1986). However, if more than one peer make judgments especially in teams of four to eight, any misjudgments from single members tend to average out (Wiley & Gardner, 2009). So, acquiring peer opinion can help cure the problem. From this point on, self-reports or peer evaluation will be used to mean a person's self-judgment plus peers' judgments on his/her performance.

Online rating and commenting became a socially appealing phenomenon through social networks such as YouTube, Facebook, and Twitter. Most such networks are tools that rely on opinions and ratings about things or people and are gaining broad community acceptance. Today's education puts too much emphasis on faculty to assess students. In that sense, faculty claim ownership and control in the process. In response to the educational trend, constructivist learning principles should be accommodated into the assessment process, which can comfortably be accomplished by self and peer assessment (Spiller, 2009). With new technological possibilities emerging every day, peer ratings, which are considered as valid as faculty assessments (Topping, 1998; Topping, Smith, Swanson, & Elliot, 2000) can be a viable alternative to assess student engagement. Following the trend, online learning platforms like massive open online courses (MOOCs) with tremendous number of attendance (Perna et al., 2014) is a potential venue for administration of peer evaluation where assessment, can become quite a burden. Peer

evaluation already is administered widely for course assessment purposes. However, there is no study, if not few, that investigates the relationship between perceived academic engagement and peer evaluation. Therefore, the purpose of this study is to examine peer evaluation as an alternative for assessing perceived academic engagement.

Online Peer Evaluation, Teamwork and Academic Engagement

Peer evaluation is particularly meaningful in the teamwork context. Assessing the individual performance in teams has been an academic discussion for long (Cooper, MacGregor, Smith, & Robinson, 2000) as students tend to exploit group work (Levi & Cadiz, 1998). Assessment, if conducted appropriately has the potential to impact engagement and, consequently, to enhance achievement (Harlen & James, 1996; Porter & Brophy, 1988). Identifying intermediate measures of academic progress is one of the keys to success in higher education (Perna et al., 2014). But, conducting peer evaluations can be very time consuming and demanding for faculty (Willey & Gardner, 2009). In this respect, an online tool has been developed in the current study to allow students to rate peer and self-performance on a project-based learning activity. Also, the tool collected information about engagement. This study reports findings from the data collected through this online assessment tool.

Participation in teamwork means involvement in course-related activities; more involvement means a greater chance for success (Chen, Lambert, & Guidry, 2010; Frydenberg, Ainley, & Russell, 2005; Marks, 2000; Sezer & Tokcan, 2003). As students who interact with each other are more successful (Astin, 1993; Hake, 1998), teamwork becomes an essential approach towards student success. Significant enhancements were reported on factors such as engagement, attendance, and active learning based on teamwork (Esposto & Weaver, 2008). Therefore, it is possible to establish a constructive relationship between academic engagement and peer evaluation conducted through teamwork. The following sections conceptually explain the key factors of this study.

Academic Engagement

Engagement is a complex concept interconnected with other educational constructs. There are several definitions and dimensions of engagement ranging from national level to international level, as well as to studies at local level (Trowler, 2010). Krause and Coates (2008), for example, define engagement as "the extent to which students are engaging in activities that higher education research has shown to be linked with high-quality

learning outcomes" (p. 493). It is also defined as the "student's psychological investment in and effort directed toward learning, understanding, or mastering the knowledge, skills, or crafts that academic work is intended to promote" (Newmann, 1992, p. 12). Kuh (2002) links engagement with seven principles for good practice in higher education (Chickering & Gamson, 1987).

There are three types of engagement: Behavioral, cognitive, and emotional. This study focuses on treats that can directly be observed and therefore deals with *behavioral engagement*. "Students who are behaviorally engaged would typically comply with behavioral norms, such as attendance and involvement, and would demonstrate the absence of disruptive or negative behavior" (Trowler, 2010, p. 5). Behavioral engagement also is organized under numerous factors such as active and collaborative learning, participation, attendance, student-faculty interaction, academic effort, preparation for class, and so forth. For the purposes of this study, types of behavioral engagement that can be directly associated with classroom activities were considered – the ones a student can self-report and peers can confirm. In this respect, two types of behavioral engagement were operationalized: (1) *active learning* and (2) *paying attention.*

Active Learning

Active learning is vigorous involvement in educational activities through behaviors such as asking questions and participating in discussions. Active learning and collaboration are major ways of behavioral engagement (Kuh, Gonyea, and Palmer, 2001). In the current study, activities such as responding to instructor questions, and participating in or contributing to class discussions (Carini, et al., 2006; Kuh, et al., 2001; Roadrangka & Yeany, 1985; Tinio, 2009) were considered to determine active learning. Active learning in this sense involves more extrovert and interactive activities in nature, compared to paying attention, which is defined below.

Paying Attention

Paying attention is focusing on and immersing oneself in educational activities through behaviors such as attending classes and perusing the class materials. Student attention is considered to be one of the foci of engagement (Trowler, 2010). Behaviors such as class contact, attendance, taking notes, and working on class materials (Carini, et al., 2006; James, Krause, & Jennings, 2010; Krause, 2005; Kuh, et al., 2001; Roadrangka & Yeany, 1985; Shernoff, Csikszentmihalyi, Schneider, & Shernoff, 2003; Tinio, 2009) were considered to determine whether a student was paying

attention. Paying attention involves activities that imply concentration and listening, that are more passive in nature, compared to active learning.

Peer Evaluation

"Peer assessment is defined as an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status" (Topping, 1998, p. 250). In general, peer evaluation is conducted through teamwork to obtain an insider's view. Students might give counterparts high scores, causing skewed results (Gueldenzoph & May, 2002; Oslin, 2010) and therefore scores can be higher than faculty ratings (De Grez, Valcke, & Roozen, 2012). Students usually do this to ensure they receive a good grade from peers or because they do not want to critique friends (Gueldenzoph & May, 2002). Nevertheless, peer evaluation is shown to correlate with grades given by instructors (Langan, et al., 2005; Lejk, Wyvill, & Farrow, 1996). There are several kinds of evaluation to assess group work. They can be synthesized into three general categories and the current study is an example of the third category: *(1) Evaluating the inclination to work in groups*: Determines whether a person is predisposed to and comfortable with working in groups in general (e.g., Rafferty, Ball, & Aiken, 2001). *(2) Evaluating a group externally*: Involves evaluating the group work typically by observation, without participating in the group activity (e.g., Morgan, Glickman, Woodward, Blaiwes, & Salas, 1986; Weller, et al., 2011). *(3) Evaluating a group internally*: Requires the evaluator to be a group participant. It is possible to differentiate two types of internal evaluation in the scope of the current study: (a) Evaluation can be done to identify issues specific to a field of study – includes concepts that are related to the respective field. For instance, evaluation on physics might require knowledge or opinions about momentum and velocity (e.g., Wiki Rubric by Franker, 2011; the Academic Report Assessment Form by Topping, et al., 2000). (b) Evaluation can inquire general group dynamics and overall impressions. Such evaluations involve statements regarding peer responsibilities and activities rather than statements specific to task content (e.g., SPARK by Freeman & McKenzie, 2002; Willey & Gardner, 2010). Lejk, et al. (1996) argue that it is easier to rate general impressions than the specific impressions, and general impressions are a reliable way of assessing peer contribution even if the rater is not a professional.

The current study is an example of evaluating a group internally – specifically the last type. The instrument developed for this study aimed to collect information from students regarding themselves and peers. Behaviors such as organizing the group work, following logic to justify decisions, and fulfilling
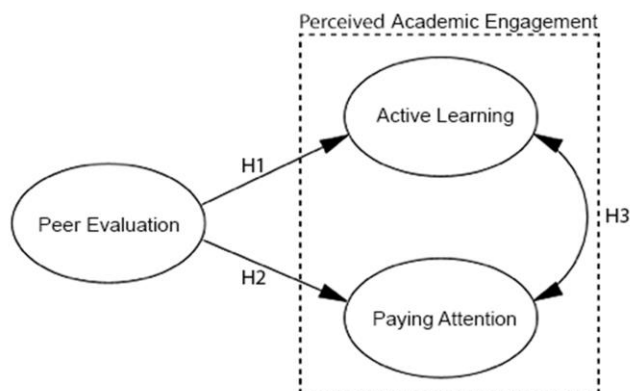
Figure 1. The research model.

responsibilities (Baker, Horvath, Campion, Offermann, & Salas, 2005; Freeman & McKenzie, 2002; Gueldenzoph & May, 2002; Lingard, 2010; Mishra, Catchpole, & McCulloch, 2009; Wang, MacCann, Zhuang, Liu, & Roberts, 2009) were considered to determine whether students performed responsibilities.

Perceived academic engagement is operationalized as a student's self-reported academic work completed in one semester. It is associated with all course activities that were performed in that semester. Peer evaluation has a narrower context, exclusive to a project work in a course, completed by a team. Consequently, the following research questions were investigated.

- How well do perceived academic engagement and peer evaluation stand as a measurement model?
- As being the two attributes of behavioral engagement, what is the correlation between active learning and paying attention; and does peer evaluation diverge enough from the former two?
- How well do peer evaluation scores represent perceived academic engagement in higher education?
- Specifically, can peer evaluation be used to predict active learning and paying attention?

METHOD

This study is an attempt to develop a model (see Figure 1) by employing SEM. Active learning and paying attention were surrounded in the figure with dashed lines. Relatively high correlation can be expected between the two constructs, as both being the indicators of student academic engagement. Students who perceive to experience active learning would also show evidence of paying attention to academic work and vice versa.

Moreover, students who receive high peer evaluation scores would report high perceived academic engagement scores.

The following hypotheses can be formulated based on the evidence suggested from the literature:

- H1: Peer evaluation will have a significant positive influence on active learning.
- H2: Peer evaluation will have a significant positive influence on paying attention.
- H3: Active learning will have a significant correlation with paying attention.

Instruments

Data were collected through an online, login-based system, which allowed students to fill out a survey and a peer evaluation form per course and per assignment. The system allowed registering multiple courses per instructor, and multiple assignments per course. Instructors could use the system for as many courses and assignments as they want. Students had to answer all questions. The data set used in this study was part of a larger set consisting of questions regarding various engagement, teamwork attitude and peer evaluation items along with demographical data. The three constructs of the current study were named active learning (ACT_LEARN), paying attention (PAY_ATTEN), and peer evaluation (PEER_EVAL). ACT_LEARN and PAY_ATTEN were originally composed of 5 items each. PEER_EVAL had 10 items. The scale items were adapted from the work of various researchers as specified in the previous sections (see Appendix A). An example ACT_LEARN item is "I participated in class discussions"; an example PAY_ATTEN item is "I concentrated in class"; and an example PEER_EVAL item is "Helped to manage disagreements and conflicts in the team". All items were

7-point Likert scale, where 1 = strongly disagree and 7 = strongly agree, except for the 10th PEER_EVAL item. For this item students typed a value between 0 and 100.

Participants and Data Collection

Data were collected through convenience sampling from three different universities during fall 2012. Including the researchers, seven instructors – offering 12 courses – voluntarily participated and invited their students to participate (see Table 1). All courses were offered in Computer Education and Instructional Technology departments on campus. Each included at least one project-based assignment. Projects in general were instructional-design related assignments, such as developing paper-based instructional materials, developing interactive educational materials, or preparing classroom presentations. Instructors registered 17 assignments into the system. Students worked in groups face-to-face to complete the

Table 1. Number of Participants by Instructor, Course, and Project.

Institution/ Instructor	Course	Invited Mid-Project	Invited Final Project	Total	Completed Mid-Project	Completed Final Project	Total	Chosen Final Project
University 1								
Instructor 1	Course 1	0	41	41	0	41	41	17
Instructor 2	Course 2	0	26	26	0	22	22	5
Instructor 3	Course 3	0	44	44	0	40	40	3
Instructor 4	Course 4	67	68	135	67	67	134	67
	Course 5	66	66	132	58	63	121	62
	Course 6	9	9	18	7	4	11	4
University 2								
Instructor 5	Course 7	0	17	17	0	2	2	2
	Course 8	0	30	30	0	1	1	1
	Course 9	0	34	34	0	4	4	4
University 3								
Instructor 6	Course 10	0	34	34	0	9	9	9
Instructor 7	Course 11	0	29	29	0	0	0	0
	Course 12	0	58	58	0	36	36	36
Total		142	456	598	132	289	421	210

Table 2. Descriptive Statistics of the Observed Variables.

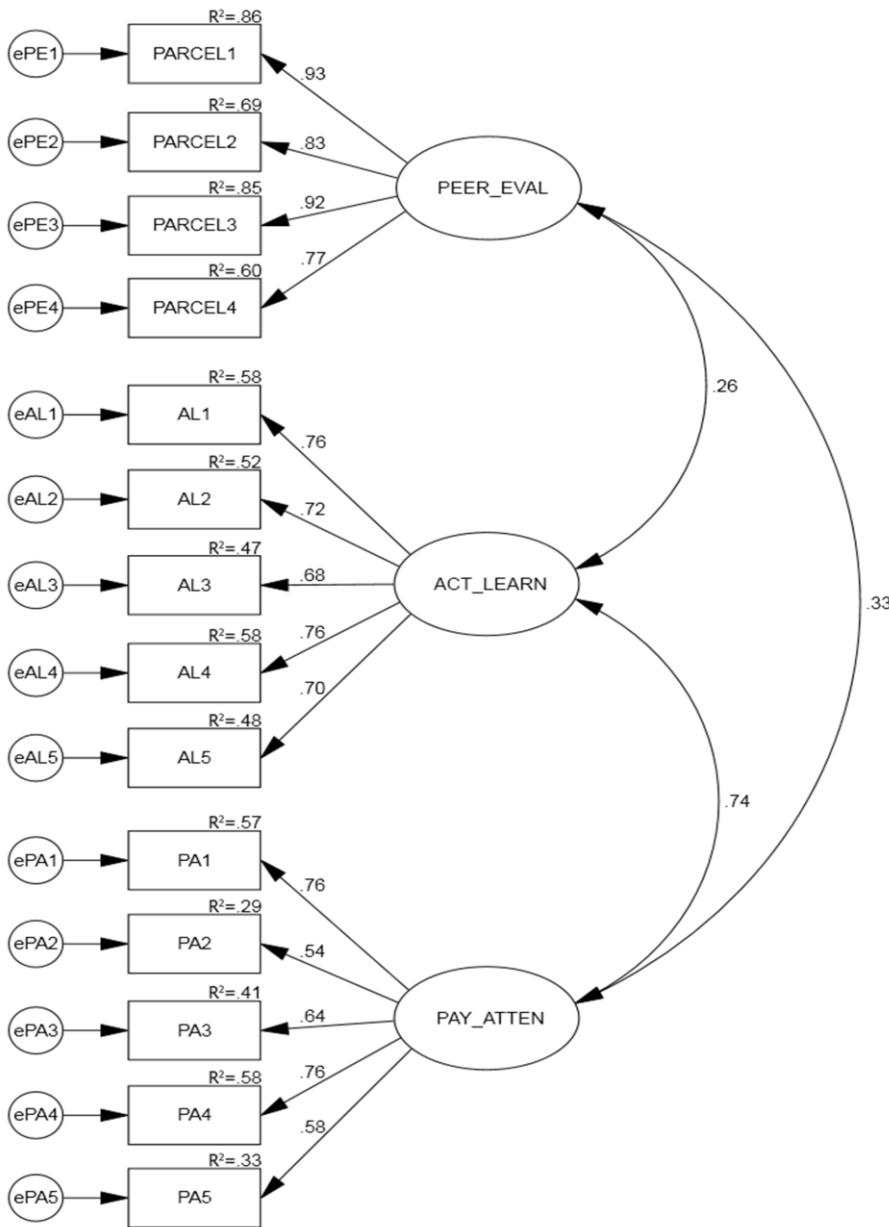|	Min	Max	Mean	SD	Skewness (SE=0.17)	Kurtosis (SE=0.33)
PEER_EVAL						
PARCEL1 (items 1 – 3)	4.00	7.00	6.44	0.54	-1.47	2.75
PARCEL2 (items 4 – 6)	4.53	7.00	6.50	0.50	-1.44	2.30
PARCEL3 (items 7 – 9)	4.40	7.00	6.56	0.48	-1.72	3.73
PARCEL4 (item 10)	5.02	7.00	6.59	0.32	-1.14	2.95
ACT_LEARN						
AL1	1.00	7.00	5.49	1.35	-0.63	-0.31
AL2	1.00	7.00	4.84	1.57	-0.65	-0.14
AL3	1.00	7.00	5.58	1.38	-1.23	1.34
AL4	1.00	7.00	5.59	1.34	-0.91	0.42
AL5	1.00	7.00	4.40	1.66	-0.21	-0.82
PAY_ATTEN						
PA1	3.00	7.00	6.04	0.97	-0.75	-0.29
PA2	1.00	7.00	5.55	1.53	-1.14	0.75
PA3	2.00	7.00	6.01	0.99	-1.10	1.39
PA4	1.00	7.00	5.85	1.12	-1.00	1.11
PA5	1.00	7.00	5.81	1.18	-1.15	1.27

Figure 2. Model 1: The measurement Model (standardized estimates). See Appendix for legend.

Table 3. Model Fit Summaries.

Fit Index	Recommended Values	Observed Values	
		Model 1	Model 2
χ^2 (Chi-Square)	Non-significant [a, b]	135.632 ($p<0.001$)	67.075 ($p=0.065$)
χ^2/df (degrees of freedom)	≤ 3.00 [c, e]	1.833	1.315
GFI	≥ 0.90 [a, c]	0.919	0.950
AGFI	≥ 0.80 [c]	0.885	0.924
TLI	≥ 0.95 [d, e]	0.949	0.983
CFI	≥ 0.90 [c]	0.959	0.987
RMSR	≤ 0.10 [c]	0.064	0.047
RMSEA	≤ 0.06 [d]	0.063	0.039

[a] McDonald and Ho (2002), [b] Kline (2005), [c] Usluel, Aşkar, and Baş (2008), [d] Hu and Bentler (1999),
[e] Schreiber, Nora, Stage, Barlow, and King (2006)

assignments. Groups were formed on voluntary bases by the participants. 598 students were invited either through mid-term and/or final project. 234 students registered into multiple courses and/or multiple assignments. 421 instances of surveys were filled. 215 unique students filled the surveys. Some students filled surveys for more than one project. In such cases, only the projects that were evaluated at the end of fall 2012 were incorporated into the analyses. Some students filled surveys in more than one course. Such duplicate instances were also removed incorporating the latest assignment. This left 210 students for analysis – 4 freshmen, 79 sophomore, 39 juniors, 84 seniors, and 4 graduate students.

Construct reliabilities were at acceptable level. The alpha reliability (Cronbach, 1951) scores for PEER_EVAL, ACT_LEARN, and PAY_ATTEN were 0.92, 0.85, and 0.77, respectively. SEM is a form of multiple regression analysis. According to Stevens (1996), it is appropriate to run a multiple regression analysis with 15 cases per independent variable. Recommended number of cases per parameter estimate can be as low as 5 (Bentler & Chou, 1987). As there were several PEER_EVAL items, artificial parceling was done, following Tinio (2009), to compact variables (see Table 2). This was done for PEER_EVAL items 1 through 9 by calculating mean scores. This helped the number of cases per observed variable to be within the limits. Item 10 was used as a separate parcel as it was collected through a different method. This item was transformed to range between 1 and 7(see Table 2) shows the parcel distributions of PEER_EVAL items and other relevant descriptive statistics of all constructs.

All items were negatively skewed as usually observed in educational settings where students report perceived attitudes. The items were mostly leptokurtic. However, all items were within acceptable skewness and kurtosis levels (seeTable 2) for SEM as defined by Kline (2005). Another simple observation shows that mean scores of the PEER_EVAL items were all above the value of 6.40, which is relatively greater than the mean score of any item of ACT_LEARN and PAY_ATTEN observed in this data set.

RESULTS

To test the hypotheses depicted in Figure 1, Model 1 (see Figure 2) was specified in AMOS to first investigate the measurement part of the model. Maximum likelihood parameter estimation was chosen instead of the other estimation methods as the data were within the normality assumptions (Kline, 2005). The figure represents each construct with the associated observed variables. The standardized factor loadings were all relatively high. High factor loadings are considered to be indicators of good convergent validity.

The analysis resulted in the model fit parameters given in Table 3, under the Model 1 column. The values were very close to what is expected from models with good fit and most values met the requirements, but the absolute indices of model fit, Chi-Square, did not return a value greater than 0.05 ($\chi^2 = 135.632$; $df = 74$; $p<0.05$). This meant that the model was not able to explain the data in absolute terms. Therefore post-hoc modifications were conducted.

In SEM it is hardly possible to obtain a good model fit at first attempt. Usually further modifications are conducted on the initially hypothesized model after examining parameters, model fit values, and residuals to obtain a better fitting model (Schreiber, et al., 2006). As the item loadings and subscale correlations were investigated based on theory, it was seen that although AL1 loaded highly on ACT_LEARN, it had high correlations with several other items, including the items of PAY_ATTEN. PA2 had the lowest item loading on PAY_ATTEN. It also had high correlations with AL1 and other items, especially with the items of PEER_EVAL. This meant that these two indicators were not discriminant enough from the other items. Therefore they were removed from the model, instead of accounting for the empirical relationships in the model – see Sams (2005) for a similar approach. The modification indices produced a number of similar suggestions (such as adding covariance between some of the observed and/or latent variables) but the model adjustment was kept mainly to theory-based respecifications. Kline (2005) denote that "learn from your data, but your data should not be your teacher" (p. 149). After the removal, the measurement part of the new model (Model 2) was tested (Figure 3). Model 2 returned model fit parameters that were all at acceptable levels (see Table 3) ($\chi^2 = 67.075$; $df = 51$; $p>0.05$). All parameter estimates were significant. The smallest of standardized factor loadings was 0.57 and all others were above 0.66 (see Figure 3 and Table 4).

Convergent and discriminant validities of Model 2 were investigated next. Teo (2009), referring to the work of Fornell and Larcker (1981), argue that there are three indicators of convergent validity of SEMs: (1) an item level reliability value above 0.70 (Hair, Black, Babin, Anderson, & Tatham, 2006), (2) a composite reliability value above 0.70 (Nunnally & Bernstein, 1994), and (3) an average variance extracted (AVE) value above 0.50 (Segars, 1997). Table 5 shows the coefficients obtained through Model 2 redefined as a measurement model. Composite reliabilities of the model are all above 0.70. Except for PAY_ATTEN, the factor loadings and AVE all returned good levels of convergent validity. Moreover, Teo (2009) states that discriminant validity for a construct is determined after comparing the square root of AVE to the correlation coefficients of that construct with other constructs. The square root of

AVE value that is larger than the correlation coefficients with other constructs for a given latent variable shows good discriminant validity. In the current study, the results of PEER_EVAL appeared to show good discriminant validity while the results for ACT_LEARN and PAY_ATTEN were very close to the assumptions.
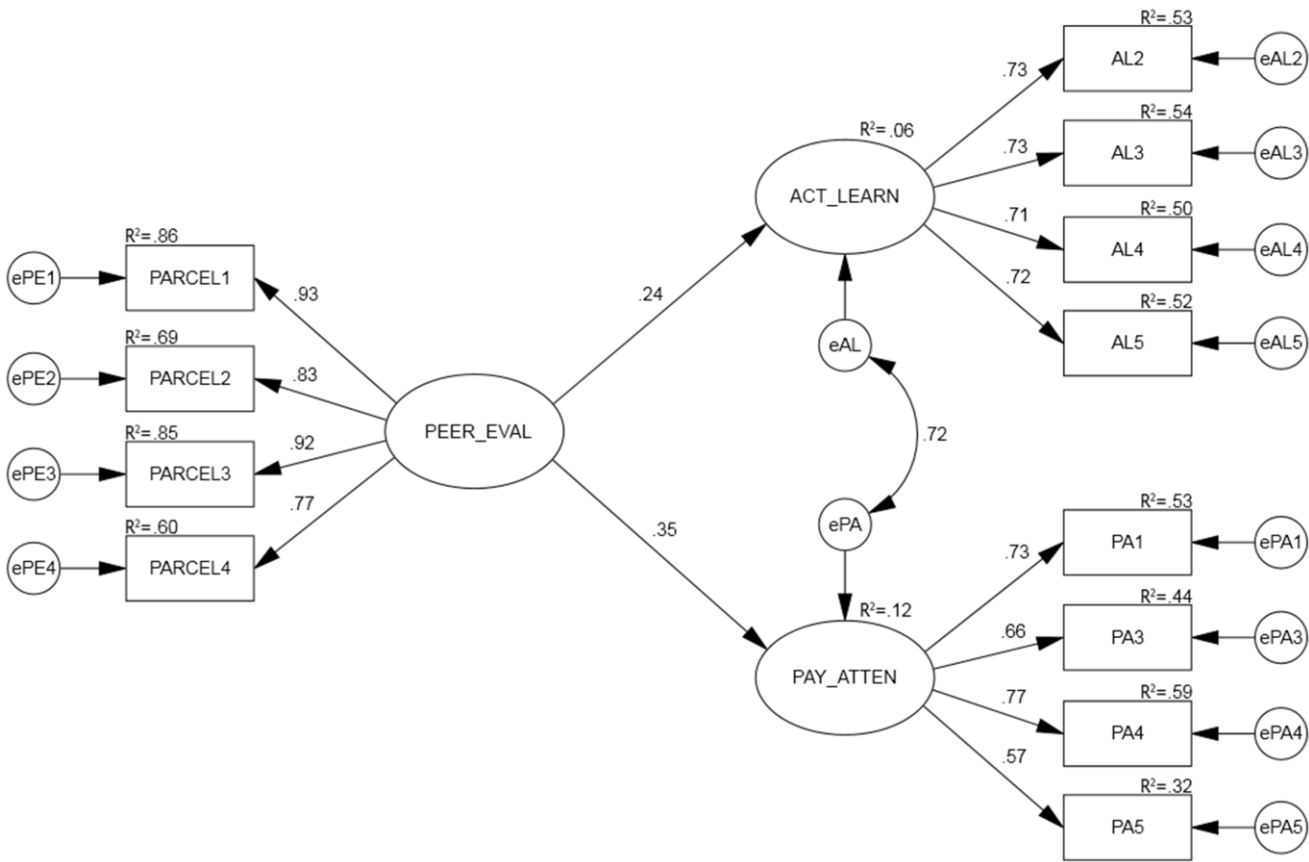

Figure 3. Model 1: The measurement Model (standardized estimates). See Appendix for legend.

Table 4. Standardized and Unstandardized Path Coefficients for Model 2.

Path[a]			β	B	SE	CR[b]	p[c]
PEER_EVAL	$\longrightarrow$	ACT_LEARN	0.24	0.54	0.18	3.00	0.003
PEER_EVAL	$\longrightarrow$	PAY_ATTEN	0.35	0.46	0.11	4.24	***
ePA	$\longleftrightarrow$	eAL	0.72	0.49	0.09	5.67	***
PEER_EVAL	$\longrightarrow$	PARCEL1	0.93	1.00			
PEER_EVAL	$\longrightarrow$	PARCEL2	0.83	0.83	0.05	17.30	***
PEER_EVAL	$\longrightarrow$	PARCEL3	0.92	0.87	0.04	21.84	***
PEER_EVAL	$\longrightarrow$	PARCEL4	0.77	0.49	0.03	15.02	***
ACT_LEARN	$\longrightarrow$	AL2	0.73	1.00			
ACT_LEARN	$\longrightarrow$	AL3	0.73	0.88	0.09	9.50	***
ACT_LEARN	$\longrightarrow$	AL4	0.71	0.83	0.09	8.87	***
ACT_LEARN	$\longrightarrow$	AL5	0.72	1.04	0.11	9.43	***
PAY_ATTEN	$\longrightarrow$	PA1	0.73	1.08	0.13	8.14	***
PAY_ATTEN	$\longrightarrow$	PA3	0.66	1.00			
PAY_ATTEN	$\longrightarrow$	PA4	0.77	1.31	0.15	8.94	***
PAY_ATTEN	$\longrightarrow$	PA5	0.57	1.02	0.15	6.78	***

[a] $\longrightarrow$ regression direction; $\longleftrightarrow$ correlation/covariance
[b] *CR* = Critical Ratio;
[c] *** *p*<0.001;

Table 5. Results for the Measurement Part of Model 2.

Latent Variable	Factor Loading (>0.70)[a]	Composite Reliability (>0.70)[a]	Average Variance Extracted (>0.50)[a]	Correlations[b]		
				PEER_EVAL	ACT_LEARN	PAY_ATTEN
PEER_EVAL		0.92	0.75	(0.86)		
PARCEL1	0.93					
PARCEL2	0.83					
PARCEL3	0.92					
PARCEL4	0.77					
ACT_LEARN		0.81	0.52	0.24	(0.72)	
AL2	0.73					
AL3	0.73					
AL4	0.71					
AL5	0.72					
PAY_ATTEN		0.78	0.47	0.35	0.74	(0.68)
PA1	0.73					
PA3	0.57					
PA4	0.66					
PA5	0.77					

[a] Recommended values for reliability or validity (Teo, 2009).

[b] Diagonals in parentheses are square roots of average variance extracted; off-diagonals are correlations between latent variables based on AMOS output.

SEMs can be viewed as confirmatory factor analyses showing causal relationships. Kline (2005) indicates that convergent and discriminant validities are accurately achieved when each observed variable loads on only a single factor in CFAs with multiple factors. Also, it is not recommended to conduct further modifications on a model once a good model fit is achieved, as the modifications might be adjusting to the attributes of the sample rather than the specifications of the theory (MacCallum, Roznowski, & Necowitz, 1992). Therefore, the analyses were ceased at Model 2. The model fit results indicate that the model has moderate convergent and discriminant validities.

The measurement portion of the model appears to stand well, with the lowest two R^2 values being 0.32 and 0.44, belonging to the items of PAY_ATTEN. All other R^2 values of the measurement portion were above 0.50, which is a reasonable number in behavioral sciences. So, the measurement part of the model is accounting for the large proportion of the variance in the observed variables. The latent variables estimate at least 32% of the variance in the observed variables. The structural portion of the model shows that there is a moderate positive connection between PEER_EVAL and engagement factors. Therefore the first two hypothesis of the study (H1 and H2) were supported. The results also confirm that there is a strong relationship between ACT_LEARN and PAY_ATTEN (0.72). H3 was also supported. This is an expected result as both latent constructs were specified to measure behavioral engagement. R^2 value for ACT_LEARN was 0.06 and

for PAY_ATTEN was 0.12. The model has less explanatory power for the variance in the latent variables compared to the observed variables.

DISCUSSION

This study was an attempt to identify the relationship between peer evaluation and perceived academic engagement in higher education. To accomplish this task, the SEM technique was utilized to stepwise test the research questions. Two models were tested. The initial model, Model 1, did not return a good model fit based on the Chi-Square statistics, even though the scale reliability results show that the survey items satisfactorily stood together. Chi-Square statistics are considered to explain the data in absolute terms. Perhaps the two items that had to be removed after Model 1 – and were originally thought to represent ACT_LEARN and PAY_ATTEN – were not uniquely characterizing the respective constructs. In other words, those two items were representative of both of the engagement constructs and possibly even of the PEER_EVAL. More rigorous results were sought by respecifications to better explain the relationship between the constructs. In this respect, advancing from Model 1 to Model 2 immediately moved the model fit results. All three hypotheses of the study were supported by the latter model. Based on this outcome, the remainder of the discussion continues on Model 2.

Peer ratings were rather high, and were skewed towards the positive end, following the literature

(Gueldenzoph & May, 2002; Oslin, 2010). The overall mean scores ranged from 6.44 to 6.59 with minimal standard deviations, corresponding to leptokurtic distribution (see Table 2). If the scale were to be considered between 0 and 100, this would translate into 90.67 to 93.17 percent. Also, the variation among ratings were lower than the variation in the engagement constructs. One way to interpret this is that students were having optimal experience in groups. Another would be that students were overrating the performance. Student ratings appear to follow the arguments of Gueldenzoph and May (2002) in that students inflated scores to receive better scores in return or because they did not want to criticize peers. This opinion gains a stronger ground especially when PEER_EVAL is compared to ACT_LEARN and PAY_ATTEN. The latter two inquired the self-judgment of engagement and were considerably lower than the former on average.

One might argue that because the student scores were inflated, whether it would still make sense to use the scores as a means to predict student engagement. The aforementioned model-fit-statistics convince us to trust the prediction. Although the overall level of scores holds substantial meaning about the performance of students, it is the variation among the scores that has the ultimate value to bring about the breadth. The results indicate that there was sufficient variation among peer evaluation scores to make distinctions about the student performances. Yet, it would be wise to seek ways that would capture better score distributions for the peer evaluation scores. Future studies can benefit from incorporating rubrics to train students on how to rate certain behaviors. Those rubrics should not be lengthy though – creating overburden, making it difficult for students to rate several team members. Informing students at the beginning of the term about the course requirements may eliminate avoidance of criticism by peers. Another solution can be to require students to justify their ratings in certain ways. Not being able to justify the given scores can build pressure on the rater, possibly yielding positive impact on the outcome. Such justifications can, then, even serve a better purpose: what Willey & Gardner (2010) emphasize as feedback. Faculty can turn these justifications to feedback given to the evaluated student, possibly in the blind review fashion.

While ACT_LEARN and PAY_ATTEN were lower than PEER_EVAL, they were higher than other benchmarks (Marks, 2000; NSSE, 2006). ACT_LEARN scores ranged between 4.40 and 5.59; and PAY_ATTEN scores ranged between 5.81 and 6.04 (see Table 2). If interpreted percentagewise, these, again, translate into 56.67 and 76.50 for ACT_LEARN; and 80.17 and 84.00 for PAY_ATTEN. Most previous engagement scores cumulate around center. ACT_LEARN scores were lower and more scattered than the PAY_ATTEN scores. The engagement scores were obtained after a teamwork activity. The referred benchmarks, however, do not necessarily involve direct teamwork activities. Engagement is a natural requirement for an efficient teamwork. So, the student ratings of engagement might have been inevitably influenced by the project work, through which the students have completed the whole survey. This does not mean that the Model specified in this study requires teamwork to work. The scores indicate that students were paying attention to academic activities more so than they actively participated. It would be logical to argue that ACT_LEARN is more demanding in comparison to PAY_ATTEN, and in this respect, observing lower ACT_LEARN scores appears rational. It is also logical to argue that ACT_LEARN, with its interactive and more demanding nature, would represent a higher level of engagement. Therefore, practitioners seeking stronger student engagement should seek ways to improve ACT_LEARN scores.

Model 2 explains that receiving high scores from peers means that the student was overall engaged throughout the semester. This means that a project work, with a partial time span compared to the whole semester, can give enough opportunities to students to identify the level of engagement that is more associated with the whole semester. That is, the specified model indeed supported the expectations. One standard deviation increase in PEER_EVAL score translated into 0.24 standard deviations increase in ACT_LEARN and 0.35 standard deviations increase in PAY_ATTEN. Consequently, the relationship between PEER_EVAL and PAY_ATTEN was 1.46 times more powerful than the relationship between PEER_EVAL and ACT_LEARN. These results support the claim that students' behavioral engagement can be predicted from PEER_EVAL. When all other values are kept constant at average, a one point increase in PEER_EVAL score meant 0.54 points increase in ACT_LEARN score on a scale where 1 = strongly disagree and 7 = strongly agree. Similarly, for an average student, a one point increase in PEER_EVAL score meant 0.46 points increase in PAY_ATTEN score. Therefore, PEER_EVAL scores have a profound effect on perceived academic engagement scores.

Online learning platforms like MOOCs, which are increasingly collaborative in nature, are getting popular, but online completion rates are low. Much drop out is attributed to the failure of courses to engage students (Baxter, 2012). Peer evaluation can be solution to the problem not only by measuring the level of student progress but also to create a means for students to better engage. Intermediary check points are suggested to keep students on track; available research does not provide much understanding on the progress of course,

except for registration and completion (Perna et al., 2014). Peer evaluation can help create intermediary check points by peer feedback without the need for extended faculty effort. However, in order to manage the peer evaluation one should place and sort scores provided by students one by one in a table format where all students are presented with the matching scores received from peers. Manually conducting peer evaluations can be a serious burden on the faculty side (Willey & Gardner, 2009). Simply encouraging faculty to use peer evaluations would not be effective. Making policy changes superficially would not make much impact, either. Institutions, interested in benefiting from the power of peer evaluation should explore ways to incorporate student peer evaluation into the systems such as online course management, online learning management, and automation (registrar). Integrating (linking) the system with social sites such as Facebook and Twitter, where people already voluntarily make comments and reviews, can facilitate the process. Such a system should not only be easy to use for instructors but also for students to be able to make a significant impact. A system like the one developed for the current study or the ones like SPARK^PLUS (Willey & Gardner, 2010) should make a substantial difference on whether peer evaluations practically sustain.

PAY_ATTEN correlated with ACT_LEARN, conforming to being the indicators of behavioral engagement. Although this was a hypothesized relationship (H3), it was stronger than what was expected in the study. The results indicate that the distinction between the two constructs were relatively low. The recommended AVE value is 0.50 (Teo, 2009). What was obtained for PAY_ATTEN is close to but lower than that. Also, it is suggested that the square root of AVE of a latent variable should be lower than the correlation between that latent variable and any other latent variable. The respective value of PAY_ATTEN (0.68) – and consequently of ACT_LEARN (0.72) – designates a challenge, as well. In Model 2, each observed variable was configured to load on only single latent variable. According to Kline (2005), such configurations are considered to fulfill the convergent and discriminant validities of the latent variables. Nevertheless, there still appears to be room for improvement in selection of items for the two constructs, in order to describe the behavioral engagement. It is a challenge to design a scale for engagement, which is multidimensional and highly dynamic nature (Annetta et al., 2009). The current trend, in general, is to estimate by individual indicator items at institutional level. Future studies can continue seeking ways to bring together suitable indicators of academic engagement.

The R^2 values for the measurement part of the model were reasonable for a study of behavioral sciences (see Figure 3 and Table 4). The smallest of the whole model was 0.32 and the greatest was 0.86. The model explained at least 32% of the variance in each observed variable. The best explained variances were of the indicators of PEER_EVAL, followed by the indicators of ACT_LEARN and PAY_ATTEN. The R^2 values of the structural portion of the model were not as high as one would expect. That is, PEER_EVAL was able explain only about 6% of variance in ACT_LEARN and 12% of variance in PAY_ATTEN. Obtaining high R^2 values in human sciences is difficult and the values usually fall below 50%, but this is in the nature of human beings – they are hard to predict. The low R^2 values simply hampers making precise estimations, but the significant coefficients obtained in the current study still show how one unit of change in peer evaluation affects the amount of change in the engagement constructs. Also, the fact that peer evaluations can fill the gap for missing engagement scores caused by irresponsive students, makes them an alternative and a complementary method for assessing engagement, as well. That is, a few team members from each group can provide reliable, or at least minimal, information about the inner workings of the whole group and the individual team members' engagement levels.

Limitations

Data were collected from several different courses. This can possibly cause variation in the latent variables. It would be ideal to choose data from only one type of course by eliminating the others. However, the number of participants was not adequate to operationalize such an option. Expanding the sample size was also challenging. The context of the study required students to work in groups. A very limited number of instructors incorporate teamwork or peer evaluation into the class activities in the way that was conceptualized in the current study.

PEER_EVAL scores were obtained through a project work that specifically asked students' contributions to that project. The engagement scores, on the other hand, asked student self-opinions about the whole semester. It would be best if the project work was a semester-wide study, matching the span of the claims students make. As explained previously, coming up with such settings is challenging. This is a potential reason for the unexplained variance in the predicted scores. Nevertheless, the relationship between the predictor and predicted was still significant.

CONCLUSION

This study established an SEM model, which fit to the data well, to predict perceived academic engagement

constructs of ACT_LEARN and PAY_ATTEN from peer evaluation (PEER_EVAL). The results showed that peer evaluation is a significant indicator of perceived academic engagement in higher education as operationalized in the current study. As teamwork and social online interaction in educational settings gain more and more emphasis, possibilities of assessment and evaluation widens. While academic engagement is considered to be a part of institutional performance (Kuh, 2002), self-reports relying only on students' own performance can mislead (Podsakoff & Organ, 1986). The current results suggest that educators and institutions can as well rely on peer evaluation scores to gain better understanding of perceived academic engagement. This requires creating more opportunities for teamwork in the academic environment and efficient ways to incorporate peer evaluation. Social networks and online learning management systems can perhaps offer viable venue to implement such necessities with the advancement of integration technologies. Findings confirm that the linkage between achievement and engagement (Carini, Kuh, & Klein, 2006; Johnson & Butts, 1983) can also be established between achievement and peer evaluation through perceived academic engagement. The claim that engaged students are successful can be succcessfuly observed and reported by peers.

Prospective research, apart from the implications and suggestions presented in the Discussion section, can possibly concentrate on improving the measurement part of the behavioral engagement constructs by incorporating new or enhanced indicators. Also, incorporating other relevant factors, such as the student characteristics, can possibly improve the variance explained by the model. For example, it is known that engagement differs based on age (Chen, Gonyea, & Kuh, 2008). The model might respond differently, if it is narrowed down to specific grade levels and genders. Moreover, in the current study, peer and self-scores were averaged together, based on the discussion made by Lejk, et al. (1996). Differentiating between peer scores and self-scores can prove useful.

Acknowledgment

REFERENCES

Annetta, L. A., Minogue, J., Holmes, S. Y., & Cheng, M.-T. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education, 53*(1), 74-85.

Astin, A. W. (1993). *What matters in college? For critical years revisited.* San Francisco, CA: Jossey-Bass.

Baker, D. P., Horvath, L., Campion, M., Offermann, L., & Salas, E. (2005). Chapter 7: The ALL Teamwork Framework. In T. S. Murray, Y. Clermont & M. Binkley (Eds.), *International Adult Literacy Survey Measuring Adult Literacy and Life Skills: New Frameworks for Assessment - Catalogue no. 89-552-MIE, no. 13* (pp. 229-272). Ottawa, Ontario, Canada: Statistics Canada.

Baxter, J. (2012). Who am I and what keeps me going? Profiling the distance learning student in higher education. *International Review of Research in Open and Distance Learning, 13*(4), 107–129.

Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods and Research, 16*(1), 78-117.

Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student engagement and student learning: Testing the linkages. *Research in Higher Education, 47*(1), 1-32.

Chen, P., Gonyea, R., & Kuh, G. (2008). Learning at a distance: Engaged or not. *Innovate: Journal of Online Education, 4*(3), 1-8.

Chen, P.-S. D., Lambert, A. D., & Guidry, K. R. (2010). Engaging online learners: The impact of web-based learning technology on college student engagement. *Computers & Education, 54*(4), 1222-1232.

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *AAHE Bulletin, 39*(7), 3-7.

Coates, H. (2005). The value of student engagement for higher education quality assurance. *Quality in Higher Education, 11*(1), 25-36.

Cooper, J. L., MacGregor, J., Smith, K. A., & Robinson, P. (2000). Implementing small group instruction: Insights from successful practitioners. In J. MacGregor, J. L. Cooper, K. A. Smith. & P. Robinson (Eds.), *Strategies for energizing large classes: From small groups to learning communities* (Vol. 2000, pp. 63-76). San Francisco, CA: Jossey-Bass.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

De Grez, L., Valcke, M., & Roozen, I. (2012). How effective are self-and peer assessment of oral presentation skills compared with teachers' assessments? *Active Learning in Higher Education, 13*(2), 129-142.

Esposto, A., & Weaver, D. (2008). Strengthening teamwork to improve student engagement in managerial economics and strategy. In *Higher Education Research and Development Society of Australasia (HERDSA).* Rotorua, New Zealand.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research, 48*, 39-50.

Franker, K. (2011). Wiki Rubric. Retrieved April 4, 2013 from World Wide Web http://www2.uwstout.edu/content/profdev/rubrics/wikirubric.pdf.

Freeman, M., & McKenzie, J. (2002). SPARK, a confidential web–based template for self and peer assessment of student teamwork: Benefits of evaluating across

different subjects. *British Journal of Educational Technology, 33*(5), 551-569.

Frydenberg, E., Ainley, M., & Russell, V. J. (2005). Schooling issues digest: Student motivation and engagement. In (Vol. 2, pp. 1-16): Australian Government, Department of Education, Science and Training.

Gueldenzoph, L. E., & May, G. L. (2002). Collaborative peer evaluation: Best practices for group member assessments. *Business Communication Quarterly, 65*(1), 9-20.

Hair, J. F., Jr., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). New Jersey: Prentice-Hall International.

Hake, R. (1998). Interactive-engagement vs. traditional methods: A 6,000-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74.

Harlen, W., & James, M. (1996). Creating a positive impact of assessment on learning. In *The Annual Meeting of the American Educational Research Association*. New York, NY.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

James, R., Krause, K.-L., & Jennings, C. (2010). The first year experience in Australian universities: Findings from 1994 to 2009. In: Centre for the Study of Higher Education, The University of Melbourne and Griffith Institute for Higher Education, Griffith University.

Johnson, T. F., & Butts, D. P. (1983). The relationship among college science student achievement, engaged time, and personal characteristics. *Journal of Research in Science Teaching, 20*(4), 357-366.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: The Guilford Press.

Krause, K.-L. (2005). Engaged, inert or otherwise occupied? Deconstructing the 21st century undergraduate student. In *James Cook University Symposium, Sharing Scholarship in Learning and Teaching: Engaging Students*. Townsville/Cairns, Queensland.

Krause, K.-L., & Coates, H. (2008). Students' engagement in first-year university. *Assessment & Evaluation in Higher Education, 33*(5), 493-505.

Kuh, G. D. (2002). The National Survey of Student Engagement: Conceptual framework and overview of psychometric properties. In. Bloomington, IN: Indiana University Center for Postsecondary Research.

Kuh, G. D., Gonyea, R. M., & Palmer, M. (2001). The disengaged commuter student: Fact or fiction? *Commuter Perspectives, 27*(1), 2-5.

Langan, A. M., Wheater, C. P., Shaw, E. M., Haines, B. J., Cullen, W. R., Boyle, J. C., Penney, D., Oldekop, J. A., Ashcroft, C., & Lockey, L. (2005). Peer assessment of oral presentations: effects of student gender, university affiliation and participation in the development of assessment criteria. *Assessment & Evaluation in Higher Education, 30*(1), 21-34.

Lejk, M., Wyvill, M., & Farrow, S. (1996). A survey of methods of deriving individual grades from group assessments. *Assessment & Evaluation in Higher Education, 21*(3), 267-280.

Levi, D., & Cadiz, D. (1998). Evaluating team work on student projects: The use of behaviorally anchored scales to evaluate student performance. In. Cal Poly, San Luis Obispo, CA: Psychology and Human Development Department, US Department of Education.

Lingard, R. W. (2010). Teaching and assessing teamwork skills in engineering and computer science. *Journal of Systemics, Cybernetics and Informatics, 8*(1), 34-37.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychological bulletin, 111*(3), 490-504.

Macdonald, J. (2004). Developing competent e-learners: the role of assessment. In *Assessment & Evaluation in Higher Education* (Vol. 29, pp. 215-226): Carfax Publishing Company.

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal, 37*(1), 153-184.

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological methods, 7*(1), 64-82.

Mishra, A., Catchpole, K., & McCulloch, P. (2009). The Oxford NOTECHS System: Reliability and validity of a tool for measuring teamwork behaviour in the operating theatre. *Quality and Safety in Health Care, 18*(2), 104-108.

Morgan, B. B., Glickman, A. S., Woodward, E. A., Blaiwes, A. S., & Salas, E. (1986). Measurement of team behaviors in a Navy environment. In. Orlando, FL: Naval Training Systems Center, Human Factors Division.

National Survey of Student Engagement. (2006). Engaged learning: Fostering success for all students In. Bloomington, IN: Indiana University Center for Postsecondary Research, School of Education.

Newmann, F. M. (1992). Student engagement and achievement in American Secondary Schools. In. New York, NY: Teachers College, Columbia University.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill, Inc.

Oncu, S., & Cakir, H. (2011). Research in online learning environments: Priorities and methodologies. *Computers & Education, 57*(1), 1098-1108.

Oslin, J. R. (2010). *Identifying programmer ability using peer evaluation: An empirical investigation*. Unpublished M.Sc. thesis, The University of Alabama Tuscaloosa, Alabama.

Perna, L. W., Ruby, A., Boruch, R. F., Wang, N., Scull, J., Ahmad, S., & Evans, C. (2014). Moving through MOOCs: Understanding the progression of users in Massive Open Online Courses. *Educational Researcher, 43*(9), 421-432, doi: 10.3102/0013189X14562423.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management, 12*(4), 531-544.

Porter, A. C., & Brophy, J. (1988). Synthesis of research on good teaching: Insights from the work of the institute for research on teaching. *Educational Leadership, 45*(8), 74-85.

Rafferty, A., Ball, J., & Aiken, L. (2001). Are teamwork and professional autonomy compatible, and do they result in

improved hospital care? *Quality in Health Care, 10*(suppl 2), ii32-ii37.

Roadrangka, V., & Yeany, R. H. (1985). A study of the relationship among type and quality of implementation of science teaching strategy, student formal reasoning ability, and student engagement. *Journal of Research in Science Teaching, 22*(8), 743-760.

Sams, D. (2005). *An empirical examination of job stress and management of emotionally-based behavior: Frontline social service personnel perspective.* Unpublished Ph.D. dissertation, University of South Florida.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323-338.

Segars, A. H. (1997). Assessing the unidimensionality of measurement: a paradigm and illustration within the context of information systems research. *Omega International Journal of Management Science, 25*(1), 107-121.

Sezer, A., & Tokcan, H. (2003). İşbirliğine dayalı öğrenmenin coğrafya dersinde akademik başarı üzerine etkisi. *Gazi Eğitim Fakültesi Dergisi, 23*(3), 227-242.

Shernoff, D. J., Csikszentmihalyi, M., Schneider, B., & Shernoff, E. S. (2003). Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly, 18*(2), 158-176.

Spiller, D. (2009). Assessment matters: Self-assessment and peer assessment. *Teaching Development Unit, The University of Waikato*: Hamilton, New Zealand.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences.* Mahwah, NJ: Lawrence Erlbaum Publishers.

Teo, T. (2009). Modelling technology acceptance in education: A study of pre-service teachers. *Computers & Education, 52*(2), 302-312.

Tinio, M. F. O. (2009). Academic engagement scale for grade school students. *Assessment Handbook, 2*, 64-75.

Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249-276.

Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment & Evaluation in Higher Education, 25*(2), 149-169.

Trowler, V. (2010). Student engagement literature review. *York: Higher Education Academy.*

Usluel, Y. K., Aşkar, P., & Baş, T. (2008). A structural equation model for ICT usage in higher education. *Educational Technology & Society, 11*(2), 262-273.

Wang, L., MacCann, C., Zhuang, X., Liu, O. L., & Roberts, R. D. (2009). Assessing teamwork and collaboration in high school students: A multimethod approach. *Canadian Journal of School Psychology, 24*(2), 108-124.

Weller, J., Frengley, R., Torrie, J., Shulruf, B., Jolly, B., Hopley, L., Hendersdon, K., Dzendrowskyj, P., Yee, B., & Paul, A. (2011). Evaluation of an instrument to measure teamwork in multidisciplinary critical care teams. *BMJ Quality & Safety, 20*(3), 216-222.

Willey, K., & Gardner, A. (2009). *Self and peer assessment: A necessary ingredient in developing and tracking students' graduate attributes.* Paper presented at the Proceedings of the Research in Engineering Education Symposium 2009, Palm Cove, QLD.

Willey, K., & Gardner, A. (2010). Investigating the capacity of self and peer assessment activities to engage students and promote learning. *European Journal of Engineering Education, 35*(4), 429-443.

◈ ◇ ◈

Appendix: Items of the Instruments

Item	Question
Active Learning (ACT_LEARN)	
AL1	▪ I asked questions when I did not understand the lesson. [a, b]
AL2	▪ I participated in class discussions. [b]
AL3	▪ I tried to answer the questions of the teacher during discussions. [b]
AL4	▪ I approached the teacher when I had to clarify something. [b, c, d, e, f]
AL5	▪ Asked questions and contributed to discussions in class. [c, d, e]
Paying Attention (PAY_ATTEN)	
PA1	▪ I concentrated in class. [a, b, m]
PA2	▪ I took down notes. [a, b]
PA3	▪ I went to classes. [b, d, f]
PA4	▪ I listened intensively to lectures. [a, b]
PA5	▪ I helped my classmates who did not understand the lesson. [b, c, e]
Peer Evaluation (PEER_EVAL)	
PARCEL1	▪ Offered new ideas and contributed to finding solutions to problems. [g, h, i, j, k]
	▪ Coordinated the team and ensured that tasks were completed. [g, h]
	▪ Showed enthusiasm and participation. [g, h]
PARCEL2	▪ Fulfilled tasks efficiently. [g, i, l]
	▪ Helped to manage disagreements and conflicts in the team. [h, l]
	▪ Made judgments to decide what to incorporate into the project report. [h, j]
PARCEL3	▪ Contributed to the quality of assignment: provided correct information, paid attention to grammar and format, etc. [h, l]
	▪ Provided helpful feedback to peers. [k, l]
	▪ Was reliable/trustable, met the given deadlines, attended the group meetings, was on time. [h, i]
PARCEL4	▪ When you consider what you have done to fulfill your responsibilities, and for your assignment to be successful, what score would you give yourself and peers out of 100. [h, i, j, k]

Items derived from [a] Roadrangka & Yeany (1985), [b] Tinio (2009), [c] Carini, et al. (2006), [d] James et al. (2010), [e] Kuh et al. (2001), [f] Krause (2005), [g] Freeman & McKenzie (2002), [h] Gueldenzoph & May (2002), [i] Lingard (2010), [j] Mishra et al. (2009), [k] Wang et al. (2009), [l] Baker et al. (2005), [m] Shernoff et al. (2003)