Achievement Emotions in Higher Education: A Diary Study Exploring Emotions across an Assessment Event

Elizabeth R. Peterson, Gavin T. L. Brown, and Miriam C. Jun
The University of Auckland

Author Note

Elizabeth R. Peterson, School of Psychology, The University of Auckland , Gavin T. L. Brown, Faculty of Education and Social Work, The University of Auckland; Miriam C. Jun, School of Psychology, The University of Auckland. Correspondence concerning this article should be addressed to Associate Professor Gavin Brown, Faculty of Education and Social Work, The University of Auckland, Private Bag 92019, Auckland, 1142, New Zealand. Email: gt.brown@auckland.ac.nz

Financial assistance to Dr. Gavin Brown for this study was provided by the Faculty of Education Research Development Fund project (grant number 3701227) at The University of Auckland for the *Self-regulation of university learning: Understanding student conceptions of assessment across cultures*. This paper is based on research conducted by Miriam C. Jun for her Master of Science thesis in Psychology supervised by Dr Elizabeth R. Peterson (primary) and Associate Professor Gavin T. L. Brown. Advice from Professor Reinhard Pekrun, Ludwig Maximilian University, Munich, Germany in preparing this manuscript is acknowledged.

Abstract

An increasing number of emotions have been found to affect the way students learn and their academic achievement. However, little is known about how dynamic these achievement emotions (AE) are, the extent to which they vary with the assessment process, and how they relate to prior academic ability and student achievement outcomes. Our intensive longitudinal diary study with tertiary students ($N = 166$) examined their AEs across a three week assessment period (study, test and feedback week). Overall, the results indicated that emotions during the study and test week were unrelated to both GPA and test score, but the starting level of emotions during the feedback week were related to GPA and test score. The changeability of emotions were not related to either GPA or test score. Overall, AEs seem to have a meaningful relationship to achievement only once results are known. These findings expand our knowledge about the relationship between AEs, prior academic ability and achievement and how emotions change across an assessment event.

Keywords: achievement emotion, control-value theory, assessment, academic ability, academic achievement

Author version of paper published as
Peterson, E. R., Brown, G. T. L., & Jun, M. C. (2015). Achievement emotions in higher education: A diary study exploring emotions across an assessment event. *Contemporary Educational Psychology, 42,* 82-96. doi: 10.1016/j.cedpsych.2015.05.002

Achievement Emotions in Higher Education: A Diary Study Exploring Emotions across an Assessment Event

1.0 Introduction

Achievement emotions (AEs) are "emotions that are directly linked to achievement activities or achievement outcomes" (Pekrun et al., 2011, p. 37). Achievement activities include tests, assignments, and performances whether in exam conditions or as take-home tasks. Achievement outcomes include scores, grades, grade point averages, and the like. These achievement activities and outcomes evoke within students a variety of AEs. For example, enjoyment, hope, pride, anxiety, boredom and shame (Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011) are some of the emotions identified in western educational environments related to academic achievement. In oriental contexts (Buchtel, 2009; Kitayama, Mesquita, & Karasawa, 2006), additional emotions, especially those having to do with relationships and social obligations (e.g., respect, calmness, indebtedness, and friendliness), have been identified as important emotions..

This paper contributes to our understanding of AEs by examining the dynamic nature of AEs through an assessment process (i.e., studying for, taking, and receiving scores on a high-stakes mid-term test) and by linking the changing nature of emotions to prior academic performance and test achievement. The paper also explores models that attempt to integrate emotions from western and oriental approaches by seeking commonalities among the emotions. It uses a relatively novel diary study method to intensively investigate repeated self-report of emotions within an assessment context. This paper could assist in improving AE theories, as well as developing insights as to how greater self-regulation of emotions may relate to students' achievement.

1.1 Theoretical Background: Structure of AEs

Research into AEs has been dominated by Pekrun's Control-Value Theory (CVT: Pekrun, 2006) which combines principles from a number of different theories including attributional theories of achievement emotion, expectancy value approaches to emotions, theories of perceived control, and models on the effects of emotions on learning and performance (see Pekrun et al. 2011 for a review). The CVT model and related empirical studies propose that AEs have three dimensions: *valence* (positive/pleasant vs. negative /unpleasant); *activation* (activating vs. deactivating); and *object focus* (activity vs. outcome).

1.1.1. Valence. In the CVT model positive AEs include emotions such as joy, hope, pride, gratitude, contentment, relaxation, and relief; whereas, negative AEs include emotions such as anger, frustration, anxiety, shame, anger, boredom, sadness, disappointment, and hopelessness (see Pekrun, Frenzle, Goetz & Perry, 2007 for a review). The same positive-negative polarity of valence has been identified in research on emotions in Japan (Kitayama, Mesquita, Karasawa, 2006) and China (Buchtel, 2009). To date the majority of research on emotion has tended to focus mostly on negative emotions (e.g., anxiety, fear, boredom, and shame), despite qualitative studies among school and university students in different academic settings (e.g., in class, studying, during tests) in which students not only reported many emotions but they described positive and negative emotions with similar frequency (Pekrun, Goetz, Titz, & Perry, 2002).

1.1.2. Effect. The effect of AEs can be considered to either contribute to greater learning effort (i.e., activating or engaging), or be neutral towards learning, or maladaptive towards learning (i.e., deactivating or disengaging). This learning effect interacts with valence so that emotions might be positive and deactivating or negative and engaging. In general, positive emotions tend to be activating except for those which seem to be associated with a sense of accomplishment that removes the need to exert further effort to learn (e.g., relaxation, contentment, and relief). In contrast, negative emotions are almost equally split

between those which activate efforts to improve (e.g., anger, frustration, anxiety, and shame) and those which are associated with decreased effort (e.g., boredom, sadness, disappointment, and hopelessness) (Pekrun et al., 2007). However, the effect of emotion on student achievement can also vary depending on the individual's cognitive resources and motivational approach. For example, a recent study that compared a group of secondary school students found that trait anxiety was negatively associated with cognitive test performance among students with lower working memory capacity; whereas, it had the opposite effect for higher working memory capacity students (Owens, Stevenson, Hadwin & Norgate, 2014)..

It is also worth noting that not all research studies agree on the effect of various AEs. For example, while there is agreement that shame is a negative, activating emotion, anger is seen by Pekrun, Goetz, Frenzel, Barchfeld, and Perry (2011) as negative activating, while Buchtel (2009) and Kitayama et al. (2006) consider it to be a negative deactivating emotion. Likewise, pride is seen as a positive activating emotion by Pekrun et al. (2011), while for the other two researchers it is a deactivating positive emotion. While these discrepancies are relatively minor, it does suggest an overall lack of agreement as to the effect of AES.

1.1.3. Object Focus. Understandably, AEs seem to be context specific within the achievement environment. AEs exist in relation to a learning activity (e.g., studying for a test) and to an outcome (e.g., getting an A or F grade on an assignment). A further context effect on AEs is whether the outcome is prospective (i.e., in anticipation of the result) or retrospective (i.e., in reaction to the actual results). Hence, by looking at the AEs in relation to both learning activities and learning outcomes, it is possible to capture the variation in AEs leading up to an assessment event (when learning and studying for and anticipating an outcome), but also during an actual assessment event (e.g., during a test), and in relation to the learning outcome (both in anticipation of the result and in appraisal of the result once it is known).

Thus, the range of AEs possible depends on the interaction among the outcome, time-frame, the valence of the emotion, and its effect. For example, a student who reports enjoying a particular science lesson as a consequence of having had a vivacious teacher could be described as experiencing a positive, retrospective, activating, learning or activity related emotion. In contrast, a student experiencing hopelessness about an upcoming test could be described as experiencing a negative, prospective deactivating, outcome (i.e. test) related emotion.

1.2. Stability of AEs.

AEs change across the assessment process; different intensities and types of AEs occur prior to an assessment while students are studying for a specific type of evaluation, during the execution of that evaluation, and after the evaluation is completed (Pekrun, Goetz, Perry, Kramer, Hochstadt, & Molfenter (2004). Goetz, Preckel, Pekrun, and Hall (2007) found that cognitive ability was related to the emotions felt before, during, and after a test with high ability students reporting more enjoyment while doing a test than low ability students, and low ability students reporting significantly more anxiety during the test and more anger across the entire testing process. Furthermore, in a cross-sectional study of tertiary students Pekrun, Goetz, Perry, Kramer, Hochstadt, and Molfenter (2004) found that students recalled anxiety most *before* a high-stakes exam, hope and relief *during* the taking of the exam, and relief *after* the exam. In general, positive AEs increased over the three process points, while the frequency of negative AEs 'anxiety' and 'hopelessness' decreased over the same time. More specifically, three negative AEs (i.e., anger, shame and sadness) increased *during,* but decreased *after* the exam, whereas, 'disappointed' increased *during* and stayed the same *after* the exam. However, a significant disadvantage of the Pekrun et al.

(2004) study was that it used retrospective recall after the examination which is likely to be less sensitive to subtle changes in emotion experienced during the actual event itself.

This last point identifies a considerable weakness in AE research. Most data collection has happened at a single time-point, despite achievement evaluation clearly having three significant time-points (i.e., prior, during, and post). In most AE studies, the AEs have been measured at only time-point in relation to the assessment event, although different time-points have been used. For example, Pekrun et al. (2009) measured the students' AEs a day prior to a specific exam; whereas, AEs measured by Daniels and colleagues (2008, 2009) were general course-related AEs collected in the middle of the semester (approximately 4-5 months after beginning of the semester).

More recently, a longitudinal study of secondary students' AEs while studying mathematics collected data at three time points over three consecutive school terms (Ahmed, van der Werf, Kuyper & Minnaert, 2013). The students' negative activating emotion anxiety remained more or less constant, while the negative deactivating emotion boredom increased over the three time points. In contrast, the positive AEs (enjoyment and pride) decreased over the three time points. Across the three school terms, students achieved more if their (a) average enjoyment was higher, (b) average pride, anxiety, and boredom were lower, (c) enjoyment and pride changed more, and/or (d) anxiety and boredom changed less. Nonetheless, a weakness of this study is that it used a general course grade measure of achievement rather than performance on a specific test or assignment. In addition, students' AEs were measured three times each separated by a school term. This means that the study has less to tell us about the subtle variations in student emotions over a specific assessment event and how those emotions and their changes might relate to learning achievement.

1.3. Learning and achievement and AEs.

The CVT proposes that AEs are associated with different academic outcomes because of their effect on study processes and behaviours. Students' AEs directly, indirectly, and reciprocally influence a variety of factors associated with their academic achievement, such as (a) level of effort employed (Capa & Audiffren, 2009), (b) engagement (Linnenbrick-Garcia, Rogat, & Koskey, 2011), (c) interest in the subject (Ainley & Ainley, 2011), (d) motivation (Weiner, 1985), (e) achievement goals (e.g., Daniels Stupnisky, Pekrun, Haynes, Perry & Newall, 2008; Pekrun, Elliot, & Maier, 2006), (f) cognitive ability (Goetz, Preckel, Pekrun , & Hall, 2007) and (g) learning strategies (e.g., Isen, 2000).

Positive emotions are generally associated with flexible and creative learning strategies, increased attention, interest, engagement in self-regulated learning, decreased task-irrelevant thinking, and greater achievement (Ainley & Ainley, 2011; Daniels et al., 2008, 2009; Graham et al., 2007; Lichtenfeld et al., 2012; Meinhardt & Pekrun, 2003; Pekrun et al.2009; Pekrun et al., 2002; Wolters, 2003). In contrast, negative emotions are generally associated with the use of more rigid learning strategies, reduced self-regulation of learning, increased external regulation of learning, and lower achievement (Ainley & Ainley, 2011; Boekaerts, 1993; Chapell et al., 2005; Daniels et al., 2008, 2009; Graham et al., 2007; Hembree, 1988; Isen, 2000; Isen, Daubman, & Nowicki, 1987; Larson, 1989; Lichtenfeld et al., 2012; Linnenbrink-Garcia et al., 2011; Nett, Goetz, & Hall, 2011; Pekrun et al., 2004; Pekrun et al., 2009; Pekrun, Goetz, Daniels, Stupnisky, & Perry, 2010; Seipp, 1991; Wolters, 2003; Zeidner, 2007).

Prior academic ability is also thought to influence AEs, potentially affecting students' sense of control over a forthcoming assessment event. For example, in anticipation of a result from a high stakes assessment, high ability students might experience relatively more hope and low ability students might experience relatively more anxiety. Similarly high and low ability students may respond differently to the actual result (e.g., contentment vs. disappointment) (Weiner, 1985; Pekrun et al., 2007). Although to some extent this will

depend on the attributions they attribute to that success or failure and the value they place on it (see Schutz and Davis, 2000; and the Appraisal component of the CVT model, Pekrun, 2006).

1.4. Range and Scope of AEs

The research described above highlights the range of research linking AEs to learning and studying behaviours and academic achievement. However, on closer examination relatively few AEs have been investigated extensively. Concern over the focus primarily on negative emotions and the limited range of emotions led Pekrun et al. (2002) to develop the now popular Academic Emotions Questionnaire (AEQ) which focuses on 9 emotions (4 positive, 5 negative) some of which are classified as activating and some deactivating. The emotions chosen for the AEQ were argued to be those that were most important within academic settings and most relevant to learning and achievement. They also chose items that tapped into the affective, cognitive, motivational, and physiological components of emotions and the three dimensions of AEs (valence, activation, and object focus) (Pekrun et al., 2011).

However, cross cultural research on emotions has applied a somewhat different lens to the study of emotions and has resulted in a slightly different, but overlapping, range of emotions. For example, Kitayama, Mesquita and Karasawa (2006) looked at university students' emotions associated with social engagement and how they differed in Japan (which is argued to be a more interdependent or collectivist culture) versus the United States of America (which is argued to be a more independent or individualistic culture). They found that, in Japan, emotions that had a relational component; for example, respect (an engaging emotion) and ashamed (a disengaging emotion) were more frequently reported than emotions that were more individualistic such as pride (an engaging emotion) and frustration (a disengaging emotion). The reverse was found for the American participants. Kitayama et al.'s research included in their list of emotions, those associated with general well-being (e.g., relaxation, calmness) or a lack of well-being (e.g., depression, disgust).

More recently Buchtel (2009) explored cultural differences in emotions with respect to their impact on their social obligations to help others. Buchtel used Kitayama et al.'s (2006) socially engaging and disengaging emotions, but also added emotions that did not have a social component such as competence and annoyance. Research in higher education contexts often has to concern itself with the various and diverse cultural and ethnic backgrounds of student participants. New Zealand is a multicultural society which is reflected in its higher education populations. Thus, a wide range of AEs, drawing on research in western and oriental traditions might be necessary.

An important characteristic of Asian students from Confucian heritage cultures, specifically, is a strong emphasis the fulfilment of family obligations and the needs of others as the basis of achievement motivation (Koh, Shao, & Wang, 2009). This sense of duty was a much stronger predictor of academic achievement (i.e., GPA) for East Asian students compared to New Zealand European students in a recent survey of tertiary students (Peterson, Hamilton, & Brown, 2013). Thus, the current study included a wide range of emotions, drawing on those identified by Pekrun et al. in the AEQ, Kitayama et al (2006), and Buchtel (2009) (Table 1 provides emotions selected, their source, valence, and effect).

<<Insert Table 1 AE by Framework about here>>

1.5. Research Needs & Hypotheses

This study aimed to address the following research needs identified in the review of literature.

1.5.1. Structure of AEs. Much of the research literature has focused on the role valence (i.e., positive vs. negative) and effect (i.e., activation vs. deactivation) play in classifying emotions. A measurement model testing the grouping of AEs as either positive or negative valence is conducted for each of the three frameworks (Table 1) separately and by

aggregating all the items across the three frameworks. A measurement model testing the grouping of AEs by their activating, deactivating, or neutral effect on learning is also tested for the Kitayama and Buchtel frameworks separately.

Despite multiple collaborations around Pekrun's AEQ, the three emotion frameworks have been studied separately. Furthermore, AEs research seems to have focused on emotions more or less in isolation (e.g., boredom in Pekrun, Goetz, Daniels, Stupnisky, & Perry, 2010; anxiety in Zoller & Ben-Chaim, 1988). The conceptual overlap of AEs may not only be between research frameworks but also within lists of emotions that seem highly similar (e.g., pride, superiority, and top of the world; anxiety, fear, shame, and guilt). Rather than treat these potentially synonymous emotions in isolation, exploratory factor analysis is used to ascertain plausible latent traits among emotions. These analyses allow us to understand better the structure and stability of emotion models.

1.5.2. Dynamic Nature of AEs. Despite extensive research on AEs (e.g., Pekrun et al., 2011) and their relationship to educational outcomes (e.g., Daniels et al., 2008), comparatively less is known about how various AEs change dynamically in response to the actual assessment process (i.e., preparation for an assessment, the assessment event itself, and the provision of scores and feedback). There are only a handful of studies that have explored changes in emotions around an assessment (e.g., Folkman & Lazarus, 1985; Smith & Ellsworth, 1987). Few have examined whether the changeability of AEs is related to the assessment outcomes.

Further, most AE research has relied on retrospective accounts of the emotions students recall feeling and those recalled emotions may be quite different to actual AEs during the assessment process itself (Bolger, Davis, & Rafaeli, 2003). Rather than rely on retrospective measures or single-shot data collection, this study uses intensive repeated measures data collection to overcome memory effects, and to establish a strong link between AEs and object focus. Intensive diary data collection captures how emotions change across a *specific* assessment or achievement context (i.e., preparing for an assessment, the assessment event itself, and the release of results and feedback). This permits a close study as to the dynamic nature of AEs.

Given that there is uncertainty as to the content and difficulty of an impending assessment and given that a mid-term test is likely to be the first formal evaluation in a course of study, it seems plausible that negative AEs would increase during the study week prior to the assessment event, while positive emotions would decrease. Once the test event itself occurs, students know the content of the test and are able to make a judgment of learning concerning their performance. Thus, it is expected that positive emotions would increase, while negative emotions would decline after the assessment event itself. Upon discovering results in the feedback week, it was expected that positive and negative emotions would remain high and low respectively because all uncertainty of performance has been removed.

1.5.3. Relationship of Dynamic AEs to Achievement and Performance. No studies have examined the effect of prior academic achievement (e.g., GPA) on the dynamic nature of AEs and whether there is a systematic relationship to test performance. Nonetheless, it is expected from the general finding that not only will positive emotions be associated with higher achievement, but that higher GPA students would report relatively more positive emotions and less negative emotions. And by extension, that positive emotions would be associated with higher test scores with the inverse pattern found for negative emotions.

It was also our expectation that students who were more volatile in their emotional responses (i.e., high changeability) throughout the assessment process would be associated with students with weaker grasp of the course content and lower test scores and GPA. The logic for this assumption comes from the idea that, because accuracy in self-assessment is

greater among higher performing people (Dunning, Heath, & Suls, 2004), students who cannot accurately assess their own competence relative to the course content and difficulty would exhibit larger changes in their AEs prior to and after an assessment task.

2.0 Method

2.1 Design

The current study is a quasi-experimental, intensive longitudinal (i.e., multiple repeated measures over a short period of time) diary study that measured tertiary students' self-reported achievement emotions (AEs). The students' AEs were monitored around a high-stakes test and the release of the test feedback.

2.2 Participants and Data Preparation

The participants were recruited from an undergraduate General Education course at a large (N≈42,000), publically-funded, research-intensive university, situated in the largest metropolitan region (approximately one-third of national population) of the country. Entry to the university is selective in that students are required to have a minimum of 150 points from the best 80 credits earned in the New Zealand National Certificate of Educational Achievement Level 3 compared to the minimum entry score of 120 points used at all other universities. General Education courses are elective, introductory-level courses specifically designed for students from outside the faculty offering the course; the university requires all students to take two such courses as part of their degree. Of the 395 tertiary students taking the course, 204 students gave voluntary permission as required by university ethics requirements to use their class-assignment responses for this study. Of these volunteers, 166 (81% response rate) qualified for the study by: (a) completing each learning log within 24 hours of being prompted, (b) completing seven or more of the nine learning logs, (c) doing two or more logs before the test and all four learning logs after the test, and (d) giving consent to access their GPA. We suggest that this is a good response rate given the number of conditions each participant had to meet.

The demographic characteristics of the final sample are presented in Table 2. The final sample consisted predominantly of Asian (43%) and NZ European (36%) and mostly female (72%). The majority of the students were aged between 16 and 29 (96%) and the mean age was 20.51 (SD = 4.31). It was possible to compare the distribution of the final sample who gave permission to use their work with the balance of the class who had completed the initial course survey for eight different demographic variables. There were no statistically significant differences in distributions for age group ($\chi^2_{(2)}$=1.00, Cramer's V=.06, p=.61), ethnicity ($\chi^2_{(5)}$=9.47, Cramer's V=.17, p=.09), English spoken at home ($\chi^2_{(1)}$=0.25, Cramer's V=.03, p=.62), birth in New Zealand ($\chi^2_{(1)}$=0.56, Cramer's V=.04, p=.46), number of courses completed ($\chi^2_{(3)}$=6.36, Cramer's V=.14, p=.10), and years attending university ($\chi^2_{(5)}$=8.91, Cramer's V=.16, p=.11). There were more females ($\chi^2_{(1)}$=7.95, Cramer's V=.15, p<.01) and fewer students first in the family to go to university ($\chi^2_{(1)}$=8.03, Cramer's V=.15, p<.01) in the participant group. However, the Cramer's V value falls in the small range (Aron, Aron, & Coups, 2005) suggesting that this can be considered relatively inconsequential to the results.

After restricting participants (n=127) and non-participants (n=148) to those who had a mid-term test score, it was found that those who gave permission to use their data in this study had a higher cumulative GPA ($F_{(1,273)}$=24.21, p<.001, d=.60) and higher mid-term test ($F_{(1,273)}$=21.68, p<.001, d=.56) by both statistically significant and practically large margins. Hence, the results of this study reflect the experiences and attitudes of the more academically able students.

<<Insert Table 2 Demographic Information about here>>

2.3 Procedure

The students were asked to complete nine learning logs (see Figure 1 for sequence and timing). The students were asked to rate the extent to which they agreed and/or disagreed with each of the 32 emotion statements (Table 2) on the day that the particular learning logs were completed. One week before the test week, the first three learning logs were completed (i.e., the study week). The next three learning logs were completed on the week of the test (i.e., the test week), and the remaining three learning logs were completed on the week of the feedback release (i.e., the feedback week) Thus, the first four learning logs were completed before the test in an average interval of two to three days and one learning log was completed on the day of the test. One learning log was completed three days after the test, but before the results were released and the remaining three learning logs were completed in an average interval of two or three days once the test feedback was released (see Figure 1).

<<Insert Figure 1 about here>>

2.4 Instrument and Measures

2.4.1 Emotions chosen for this study. Emotions used in the study were drawn from three sources. The nine emotions words in the AEQ (Pekrun et al., 2011), 24 emotions from Kitayama et al. (2006), and 22 emotions from Buchtel (2009) resulted in 17 unique positive emotions and 15 unique negative emotions (see Table 1).

2.4.2 Learning log. Structured diary methods are longitudinal, repeated-measures, data collection techniques in which participants record their thoughts, emotions, and actions using identical self-reports as they experience an event (DeLongis, Hemphill, & Lehman, 1992). Each questionnaire within the diary covers the events and experiences of a brief time period (e.g., a few hours or a day) and participants are asked to complete a diary section at regular intervals (e.g., daily). More intensive sampling (e.g., every hour or so) may take place in the context of monitoring some processes (e.g., smoking cessation). The structured diary method addresses questions concerning process and change and reduces the possibility of using an aggregated response to reconstructed events and the chance of participants' current state influencing their recall (Bolger et al., 2003; DeLongis, Hemphill, & Lehman, 1992).

In the current study, in order to reduce possible memory or social desirability effects arising from an open-ended diary we used a closed-format rating scale in which participants indicated their level of agreement. To ensure responses were linked in time to a known event, all diary responses had to be completed within 24 hours of the end of the 2-hour course lecture. Reminders were sent to all students concerning the need to fill in their diary within 24 hours and no entries were accepted once the entry period expired. The online learning log (i.e., diary) was administered using 'Survey Monkey' and asked participants to indicate agreement using a positively packed rating-scale, with four positive and two negative options. Positive-packing (i.e., increased number of response options in the positive direction) is deemed appropriate when participants are inclined to agree, a situation likely in terms of academically successful students (Brown, 2004; Klockars & Yamagishi, 1988; Lam & Klockars, 1982). Furthermore, good measurement characteristics (i.e., ordered response options with relatively equal intervals) were found with this rating scale in a study of Chinese higher education students (Deneen, Brown, Bond, & Shroff, 2013).

2.4.3 Academic ability. With students' consent, their grade point averages (GPAs) as at the end of the semester in which the study was conducted were retrieved from the university database. Academic grades are recorded as letter grades (A+, A, A-, …, D) which were converted to a numeric values ranging from 9 (A+) to 0 (D+ and D).

2.4.4 Academic achievement. Academic achievement was measured using student performance on the first assessment in the course; that is, a moderately high-stakes, 50-item multiple choice mid-term test, worth 25% of the course grade. Students were not required to pass this test to pass the course, since performance on all three course assessments is summed for the final course grade. All students completed the same test which measured the

content (i.e., cognitive processing, forgetting, general learning theory, memory, meta-cognition, retrieval, and schema) taught in the first five lectures of the course. Simple classical test theory marking was used (i.e., one mark per each item answered correctly); the Cronbach alpha estimate of reliability was $\alpha = .71$, $M = 23.78$; $SD = 5.96$; $SEM = 3.21$. The test score was reported as a percentage, with 50% set as the pass mark in accordance with University standards, and was used as a continuous variable. Scores were retrieved from the University database.

2.5 Data Analytic Strategy

In accordance with Little and Rubin (2002), cases with a small amount of data (i.e., not more than 29 of the 288 repeated statements) were imputed using the expectation maximization (EM) technique within SPSS 20 (IBM, 2011).

2.5.1 Factor analysis. As per the CVT, confirmatory factor analysis measurement models were tested for the following models:

(Model 1) positive vs. negative structure of AEs for (a) Pekrun, (b) Kitayama, and (c) Buchtel frameworks,

(Model 2) positive vs. negative structure of all AEs aggregated across three frameworks, and

(Model 3) positive vs. negative structure with subordinate effect (i.e., engaging, neutral, and disengaging) factors for (a) Kitayama and (b) Buchtel frameworks.

In addition, because the AEs used in this study were taken from three different frameworks, exploratory factor analysis of the Time 1 data was used to identify the dimensionality of the emotions. Procedures outlined by Courtney (2013) identified two (Comparison Data Pearson RMSR eigenvalue =.186, $p<.001$) and three (Velicer's MAP^2 =.017) factors as most defensible. Exploratory factor analysis (maximum likelihood extraction with oblique minimisation, Costello & Osborne, 2005) rejected the three factor solution because no items loaded >.30 on the third factor. A two-factor solution recovered the positive vs. negative valence solution. This solution was tested in confirmatory factor analysis as Model 2.

Since Model 2 had 17 items in the positive factor and 15 items in the negative factor, exploratory factor analysis was run separately on the positive AEs and the negative AEs to better ascertain whether there was a conceptually-valid latent structure within each set of AE items. Items with factor loadings > .40 were kept, while items with cross loadings on another factor >.30 were excluded (Bandalos & Finny, 2010). This produced five factors (two positive AEs and three negative AEs) (Table 4). Then confirmatory factor analysis (Bandalos & Finney, 2010) using AMOS 20 program (IBM, 2011) tested separately the fit quality of the five emotion factors, followed by invariance testing across the 9 times of administration (Table 5).

2.5.2 Invariance testing. Nested multi-group confirmatory factor analysis (Cheung & Rensvold, 2002) was used to determine the stability of all measurement models (1a,b,c; 2; 3a,b; and 4a,b,c,d,e,f). The purpose of this testing is to determine whether the measurement model detected at Time 1 of the diary was statistically equivalent across the nine learning log entry days. If required, items with low loading would be removed to ensure invariance across time and items with negative error variances having critical ratio values with $p>.05$ were corrected to a small positive value (Chen, Bollen, Paxton, Curran, & Kirby, 2001).

For each factor a sequence of increasingly constrained models was tested: (a) equivalent regression weights, (b) equivalent factor intercepts, and (c) equivalent item residuals. Difference in the CFI $\leq .01$ as each parameter is constrained indicates statistically equivalent models. Conventionally, equivalent regression weights (metric equivalence) and intercepts (scalar equivalence) are needed to proceed with comparisons, although McArdle (2007) has argued that metric equivalence is sufficient to compare factor means in

longitudinal conditions. Lack of statistical equivalence at the intercept level may be consistent with the notion of item impact (Zumbo, 1999), in that non-equivalence reflects real-world differences in person responding to the inventory items due to the changing assessment conditions at the time of each survey day.

2.5.3 Longitudinal analysis. Models that met invariance testing at the metric level were further tested for longitudinal change using latent curve modelling (LCM) (Bollen & Curran, 2006). LCM presumes there is a linear relationship over time between starting (i.e., intercept) and tendency to change (i.e., slope) values in the constructs being evaluated. Two inter-correlated latent traits (i.e., starting mean and changeability) are used to explain variation in responding over time (Figure 2). The starting mean is kept constant in the model and the change value was indexed to the number of days elapsed since the start of the study.

The exploratory factor analysis models of emotions (Models 4a-f) were then examined across a 9-day LCM model for each emotion as per Figure 2. These were also compared to three week-long models. Week long models were chosen in order to try and capture changes in emotions during (a) study week (administration days one to three), (b) the test week (administration days four to six) and (c) feedback week (administration days seven to nine).

Inspection of model fit indices was used to select the better approach to analysing the data (i.e., 9 day models or 3 weekly models). As well, LCM models were used to investigate (a) the predictive effect of GPA onto the intercept and change factors and (b) the effect of intercept and change as predictors of test score. Note that in Week 1, three 3-day LCM models were tested because they all had at least metric invariance. This was done to ensure that they were admissible (i.e., all covariance matrix were positive definite and all negative error variances have low critical ratios).

<<Insert Figure 2 about here>>

2.5.4 Model quality. The quality of all CFA and LCM models was determined by comparison of model results to data. Non-rejection of a model was accepted when multiple indices indicated sufficient quality (Hu & Bentler, 1999). In line with current practice (Cheung & Rensvold 2002; Fan & Sivo 2007; Marsh Hau, & Wen, 2004; Vandenberg & Lance, 2000), criteria for excellent fit were models with statistically non-significant χ^2/df ratio (esp. <3.00), gamma hat ≥ .95, and root-mean square errors of approximation (RMSEA) and standardized root-mean residuals (SRMR) ≤.05 and .06, respectively. Acceptable fit was determined if the χ^2/df ratio was statistically non-significant (i.e., <3.83), gamma hat ≥ .90, and if both RMSEA and SRMR were ≤.08. Greater weight was put on the gamma hat and SRMR values, since these have been shown to be resistant to model complexity and mis-specification, unlike CFI and RMSEA which respond inversely to complex model or mis-specification (Fan & Sivo, 2007). Models that had gamma hat and SRMR values that met at least the acceptable thresholds were not rejected, while those which met the excellent fit standards were preferred.

Generally, LCM requires large samples to ensure accurate estimation of parameters and avoid improper solutions such as negative error variance. Even with samples as large as 400, improper solutions can arise by chance about 2% of the time (Boomsma & Hoogland, 2001). This study had a much smaller sample size, which was compensated for in the 9-day LCM by having many data points in the model. However, by breaking the 9-day model into three 3-day models there was a significant reduction in the information available to ensure proper solutions would be obtained. Negative error variances can be fixed to a small positive value (.005) if there is evidence that (a) in larger samples or more complex models the error variances are not negative and (b) the observed error is likely to be a chance artefact (i.e., the observed value is less than two times the standard error) indicating the true value is unlikely to be negative (Chen Bollen, Paxton, Curran, & Kirby, 2001). Because the 9-day LCM for

each emotion did not have negative error variance and because the observed error variances in the week-long LCM models were less than two times the standard error, three negative error variances in the shorter week-long models were corrected to the recommended value of .005. Alternatives to linear models were tested (i.e., quadratic and logistic) and these were found not to reach even marginal levels of fit for non-rejection; hence, linear models for each of the three weeks were preferred.

3.0 Results

First, the results from the various models developed to identify the structure of AEs are presented, followed by invariance testing of the factors across time. Then, latent curve modelling results are presented to show the effect of the assessment process and GPA on emotion variation and the effect of emotion upon achievement across time.

3.1 Factor Analysis and Invariance Testing

All three independent positive vs. negative valence models (Models 1a, b, c) had good fit with only the AEQ having strong invariance and the other two having metric invariance (Table 3). This finding supports the analysis of the latent factors as repeated measures across the 9 days.

Likewise the positive vs. negative valence model for *all* AE items (Model 2) had good fit but with weak invariance across the nine days. The hierarchical nested positive-negative model with subordinate factors for engaging, neutral, and disengaging had metric invariance for the Kitayama emotions and was inadmissible due to negative error variance for the Buchtel emotions (models 3a and 3b).

All of the positive and negative emotion 9 day models (Models 1a-c and 2) were tested in latent curve model analyses for the three week-long data collection period (i.e., study week, test week and feedback week and with GPA) (see Models 4a -4f). Models 4a-4c were inadmissible because of non-positive definite covariance matrices among at least one pair of correlated emotion valences. This is often an indicator that too many factors have been specified, which can arise given the low ratio of cases to variables. In contrast, Table 3 shows that Models 4d -4f (which contained all positive vs. negative emotions) was admissible using the three week structure, perhaps due to the higher ratio of cases to variables. However, the three weekly models had fit indices so low that the models should be rejected as not representing the underlying data. Additionally, the three week models had no statistically significant paths between the intercept and change values in the latent curve model and the GPA or test score.

<<Insert Table 3 Alternative model results about here>>

In contrast, the exploratory factor analysis of the positive and negative AEs separately resulted in two factors for the positive emotions and three factors for the negative emotions (Table 4). The first positive emotion, labelled ‘happy’, contained AEs drawn from Pekrun’s AEQ (i.e., enjoyment), Kitayama’s framework (i.e., friendly, respected), and Buchtel’s framework (i.e., appreciated and competent). The second positive emotion, given the working title of ‘chilled’, drew predominantly from Kitayama’s framework (i.e., AEQ: relief; Kitayama: superior, top of the world, relaxed, sympathetic, and calm). The label ‘chilled’ is suggested as appropriate since the AEs seem to point to the student feeling serene and above the assessment situation, perhaps as a consequence of mastery and high performance. The three negative emotions were ‘sad’ (i.e., sad, unhappiness, depression, and sulky—all drawn from Kitayama’s framework), ‘anxious’ (i.e., frustrated, anxious, fearful, and bored—one from AEQ and balance from Kitayama), and ‘self-loathing’ (i.e., ashamed, disgusted, angry, and guilty—one from AEQ and balance from Kitayama).

<<Insert Table 4 EFA results about here>>

Each exploratory factor with its contributing items was tested with nested multi-group confirmatory factor analysis and, after trimming of low loading items to improve model fit, it

was found that four of the five factors had statistically equivalent regression weights (i.e., metric equivalence) across all nine diary administrations (the exception being Self-loathing) (Table 5). Only Sad had strict equivalence, and the remaining three failed to reach equivalent intercepts (i.e., scalar equivalence). This indicated that the relationship of the emotion latent trait to the items was not stable across the nine times of administration, supporting the argument that the assessment process impacted responses to the stimulus items.

<<Insert Table 5 final invariance about here>>

3.2 Factor Mean Score Analysis

The factor mean scores for each administration day were calculated by finding the average rating of the contributing items for each AE factor. Overall, students rated the positive emotions higher than the negative emotions across the nine administration days (see Figure 1).

In keeping with our hypothesis, the two positive emotion factor means had similar trajectories across the nine administration days, with decreases as the test day drew nearer, large increases after the test, and gradual declines back towards baseline once the test feedback was released. Also in keeping with our hypotheses, the three negative emotions had similar patterns to each other, and inverse increases and decreases to the positive AEs. The negative activating AE 'Anxious' changed more around the test and its feedback, compared to the two other negative emotions.

3. 3 Latent Curve Modelling (LCM)

Given that the LCM based on the positive-negative valence of emotions was either inadmissible or non-fitting, respectively (see Table 3 models 4a-f), LCMs for the five emotion factors consisting of the full factorial model (i.e., latent trait and contributing items) repeated for each diary administration were used (Table 6). Table 6 show the fit of the five emotion factors in both the 9-day model and the three week-long models (study, test and feedback week). The fit of the three one week models was superior to the full 9-day model. More specifically, the 9-day models had much higher SRMR values with all values over .075 and two were beyond the acceptable threshold of .08. In contrast, the week-long models had SRMR values less than .075 (and seven of 15 models had values less than the excellent threshold of .06). Further, the week-long models also had much lower gamma hat values (the 9-day models had three less than the acceptable threshold of .90 vs. the week-long models having six values equal or above the excellent threshold of .95). While the 9-day models had superior RMSEA fit (two less than or equal to the acceptable threshold of .08 vs. just two in the week-long models), this statistic rewards the greater complexity of the 9-day models and should be discounted. All models had good (although four week-long models had values >3.00) or excellent indices for the ratio of χ^2 to *df*. The exception to the superior fit of the week-long models was the 'Feedback week' model for 'Anxious', which was unacceptable ($\chi^2/df = 5.75$). Consequently, this instance of 'Anxious' was removed from further analysis.

<<Insert Table 6 final LCM fit about here>>

3.3.1 Effect of AEs on test scores. Table 7 shows the independent effects of emotion intercepts and slopes on the students' test score across the three week assessment period. The results indicated that both the starting points and changeability of emotions during the study and test week were statistically unrelated to student performance on the test. In the feedback week the starting levels of the students' emotions were related to the students test performance, such that higher performing students on the test reported higher positive emotions and lower negative emotions compared to the students who had a low score on the test. With the exception of self-loathing in the feedback week, the fit of these models was found to be acceptable to good (See Supplementary Table 1).

<<Insert Table 7 LCM on Test Score about here>>

3.3. 2 Effect of GPA on AEs. Table 8 shows the independent effects of student GPA on the intercept (starting mean) and slope (changeability) of the five AE across the three assessment weeks. These results show that students' prior ability (GPA) has no statistically significant effect on their emotions levels or their changeability during the study or test week. However, GPA related to the starting levels of emotions in the feedback week. That is, in keeping with the test score results, students with higher GPA reported higher starting levels of positive emotion and lower starting levels of negative emotions once the test results were known. The fit of these models was found to be acceptable to good (See Supplementary Table 2).

<<Insert Table 8 about here>>

3.3.3 Effect of GPA and test score on AE. Having established that both GPA and test scores had statistically significant relationships to AEs only in the feedback week, we examined the combined influence of students' GPA on emotions and emotions on students' Test score, while taking into account the direct effect of GPA on test score. Table 9 shows that the joint models had acceptable fit and that GPA was consistently and positively a strong predictor of students' tests scores (variance explained SMC >.30). In keeping with the independent models, when GPA and test score were modelled together, higher GPA student had significantly higher starting means for happiness and significantly lower starting means for the negative emotions factors (sadness and self-loathing) in the feedback week (Table 9). However, only the starting point and changeability of students' positive emotions in the feedback week were significantly associated with higher performance on the mid-term test.

<<Insert Table 9 AE to test and GPA to AE about here>>

3.3.4 The curious case of anxiety. As shown in Table 6, the anxiety model for the Feedback week did not have a statistically significant fit to the data. Nevertheless, anxiety did have an acceptably-fitting model and statistically significant relationship to test score in the test week. In that period, changeability in anxiety had a statistically significant negative effect (β= -.45, p<.05) on test score, indicating that those whose anxiety was unstable during the test week tended to perform more poorly on the test. To further understand this we examined the mean anxiety levels of students who performed in the top and bottom quartiles of the test. Figure 3 shows that students with the highest test scores *slightly agreed* (score 3.00) that they were anxious two days before the test and this was significantly more than the lowest quartile students ($t_{(90)}$ = -2.04, p = .04). For upper quartile students, there was a more gradual decline in anxiety, compared to the lowest performing students who both *mostly disagreed* they were anxious and whose level of anxiety dropped more rapidly after the test. This shows that the lower performing students were more variable in their level of anxiety than highest performing students.

<<Insert Figure 3 about here>>

3.3.5. The curious case of boredom. Boredom initially appeared as a factor within the anxious factor but was removed from all subsequent models due to model-trimming processes to establish stable, invariant factors across nine times of administration. This is a striking, non-replication of previous studies which have identified strong effects for boredom, which is conceptually the inverse of interest as a motivating factor. There are two plausible explanations for the disappearance of boredom. First, this course is a general education course, which has to be from a pool of introductory courses from outside the student's own faculty. Thus, the role of interest or lack of it (boredom) may not play a significant role when students enrol in this course to complete compulsory general education credits. Secondly, and perhaps more importantly, the students who chose to participate in this study generally scored much better on the test than those not participating. Hence, the effect of boredom or interest would become moot in these circumstances, since able students are more likely to

exercise individual interest to compensate for a lack of situational interest in their learning (Alexander, Murphy, Woods, Duhon, & Parker, 1997).

4.0 Discussion

We first examine the mean scores for each emotion factor across the assessment process, then how those emotions relate to performance on the mid-term test and the effect of GPA on emotions, before considering the implications of the joint GPA, test score, and AE relationships. We conclude by touching on practical implications of the study, directions for future research, and the significance of this work.

4.1 Structure of AEs

We were able to successfully recover the positive-negative valence structure of the three frameworks selected for this study and demonstrate that this structure had metric invariance across the nine-days of administration. More telling though, is the lack of admissibility in the three framework LCM models of Pekrun et al. (2007), Buchtel (2009), and Kitayama et al. (2006) at Week 1 and the poor fit of the full positive-negative valence models at each week to fit the data. Combined with statistically non-significant paths to and from test score and GPA respectively, it would seem, within the constraints of this study, construing emotions as purely positive or negative is an unhelpful and inadequate approach.

In contrast, our exploratory approach identified five conceptually meaningful factors (although the terminology used to label them is open to debate). We suggest that these conceptual factor bundles provide a more insightful approach to understanding student AEs than focusing solely on the valence or effect of individual emotions. A major contribution of this study is the integration of Pekrun et al.'s (2007) AEQ research with that of Kitayama et al. (2006) and Buchtel (2009) in identifying emotion factors that draw on emotions each has identified as having merit. The emotion factors we have identified (i.e., Happy, Anxiety, Sad, and Chilled) have similarity to Pekrun's emotions of Enjoyment, Anxiety, Hopelessness, and Relief, while our factor Self-Loathing combines Pekrun's emotions Shame and Anger.

We have offered speculation above as to the disappearance of anxiety and boredom, and will not rehearse those comments here. In addition, the socially engaging and disengaging emotions used by Kitiyama et al (2006) and Buchtel (2009) that were in our study were not identified as separate factors despite the relatively high proportion of East Asian students in our sample. Indeed, the more sophisticated sub-structure of effect of emotion (i.e., activation or engagement vs. deactivation or disengagement) was recovered for Kitayama's AEs, but was inadmissible for Buchtel's framework. It may be that the current study had insufficient power for this sophisticated approach. However, we consider that the engagement effect may be of less interest than the substance of student AEs as this study has discovered.

Needless to say, the current result is very tentative, arrived at through many data cleaning and trimming steps, and is in need of independent replication. Nevertheless we successfully tested pre-existing conceptualisations of emotion valence and emotion activation and found both of these not to be warranted by the current data.

4.2 Factor Mean Change

Examination of the means of our five emotion factors across the assessment event suggest that students' emotions tend to vary according to when they were measured during the assessment process (Figure 1). In keeping with our hypotheses, and the findings of Pekrun et al (2004), our results seem to suggest that preparing for a test induces on average fewer positive AEs and slightly more negative AE, whereas finishing the test has the opposite effect. However, (with the exception of Anxiety) the changeability of these emotions in the three week models (study week, test week and feedback week) was not a significant predictor of students' outcomes, nor was it related to students' GPA. This finding is discussed further below.

4.2 Relationship between AEs and Test Scores

The relationship between AE and students test scores was different from the hypotheses. We expected that, leading up to an assessment event (study week and test week), students with high and stable positive activating emotions would have higher test scores and that those with high and unstable negative deactivating emotions would have lower test scores. However, we found that the starting levels of students 'emotions and the changeability of them leading up to an assessment event had almost no impact on students' test performance (the only exception to this was anxiety, discussed below). Schutz and Davis's (2000) model of emotion regulation during test taking, and the environment (task demand, and value) component of the CVT offer a potential explanation for this finding. It could be the moderately high stakes assessment event was 1) not of sufficient value to the students and or 2) that the assessment task was not perceived as sufficiently demanding to induce high levels of emotion leading up to the test. In turn, this level of emotion may have been insufficient to induce study behaviours that meaningfully affected test performance. Unfortunately, we did not assess the importance of the assessment to the students or the perceived demands on the test, so we are unable to verify this possibility.

In keeping with our hypotheses we did find that emotions related to test score performance once the results had been received (i.e., in the feedback week). Here we found the typical pattern of higher performing students reporting greater levels of positive emotion and lower levels of negative emotions compared to low performing students. This finding supports Pekrun's (2006) CVT which posits that happiness is induced more than other AEs following a success, and anger and sadness are induced more than other AEs following a failure. The current study also adds to the CVT, by finding that feelings of being chilled is also experienced by students following a success.

Again, Schutz and Davis's (2000) model of emotion regulation during test taking, and the appraisal (expectancy and attribution) component of the CVT offer a potential explanation for this finding. Both models suggest that students' response to the grade depends on the meaning they attach to it. Further, the meaning students attach to a grade depends on the goals and self-beliefs of the student. It seems that that the high performing students on the test in this study, attached sufficient meaning to the test to lead to an increase in the happiness and a sense of being chilled and a decrease in their levels of sadness and self-loathing.

Overall our findings in relation to emotions and test performance provide almost no evidence for the part of the CVT model that suggest that emotions influence learning, but they do provide support for parts of the CVT model which suggests that achievement outcomes are antecedents for students appraisals and their associated emotions.

4.3 Relationship between GPA and the AEs

With respect to GPA, in contrast to our hypothesis, we did not find that GPA influenced the changeability of students' emotions over an assessment event. This hypothesis was based on the CVT model which suggests that if students are unable to control emotions that are not positively related to the assessment task, it can drain their cognitive resources affecting their study behaviour and learning strategies and ultimately their achievement. Instead we found that GPA was only important for the starting levels of students' emotions once the students had received their mark in the feedback week. That is, in line with our findings for test score, high GPA students reported greater positive and fewer negative emotions once they knew their results.

These findings are broadly in line with those of Daniels et al., (2008) who reported that tertiary students' high school averages (academic ability) were not related to their level of enjoyment, boredom, and anxiety measured in a one off assessment towards the end of the tertiary academic year. They are also in line with Folkman and Lazarus (1985) who found

that GPA was not a significant predictor of emotion two days before an assessment event suggesting that emotions while studying are more influenced by present and immediate concerns than by past performance.

These findings also provide some support for the CVT model and Schutz and Davis's emotion regulation model in that the emotions felt after an assessment result become part of a feedback loop, which affects future performance and over time GPA. More specifically, both models suggest that emotions in response to an assessment outcome are related to the appraisal of that outcome and the attributions associated with those appraisals. For example, if a student performs poorly they may feel shame if they attribute the failure to a lack of ability, or guilt if they attribute the failure to a lack of effort. These emotions, in turn, can influence future study strategies as well as future expectations. Cumulative experience of perceived success or failure on a assessments, over time is believed to influences how students see themselves as learners, these beliefs may be self-fulfilling and, hence, related to students' cumulative GPA.

4.4 Relationships between GPA and AE and AE and Test score

When we modelled the relationship between GPA and emotion (starting mean and changeability), and emotion (starting mean and changeability) on test score together, we found that while a higher GPA was associated with greater positive and fewer negative emotions once the test results were known, only the positive emotions were related to students' actual test scores. This suggests that 1) the students who performed more poorly on the test may not have valued the test sufficiently to lead to a significant negative emotional response (i.e., sadness or self-loathing) and 2) overtime, students who consistently get higher grades may develop a greater overall level of well-being due to experiencing not only more positive emotions, but fewer negative emotions.

With respect to anxiety, it seems that, in our sample of students, slightly agreeing that one is anxious a few days before the test (as opposed to mostly disagreeing one is anxious) may be a good thing in terms of getting a higher test score. While meta-analyses have found that test anxiety before during and after a test is negatively related to achievement (e.g., Chapell et al.,2005; Richardson, Abraham & Bond, 2012), it is important to note that the current study was not asking specifically about test anxiety, but rather asked the students to rate emotion statements based on their experience that day (e.g. I felt happy, I felt anxious). However, our findings are in keeping with the idea that, while anxious people (compared to non-anxious people) may engage in more task-irrelevant behaviours which can use up cognitive resources and lower performance, controlled levels of anxiety may also lead individuals to exert more effort in order to overcome their anxiety. Further, provided that anxiety levels are not too high (as in the current study), and the task demands are not too high, this compensation may lead to better test performance (Eysenck, 1979). Another explanation might be that heightened anxiety two days before the test may help prevent students from becoming over-confident which might lead to reduced effort. That the levels of anxiety of the lowest and highest performing students was not significantly different on the evening after the test (see Figure 3), suggests that anxiety level affects students' preparation before the test, rather than performance on the test. However, more research is needed to unpick this finding.

4.5 Summary

The study confirmed that positive and negative AEs of both activations are experienced in an assessment context and around its feedback. It also confirmed that anticipation of the test event generally decreases students' positive AEs and increases students' negative AEs and that after the test the effect is reversed.

In general, we found that emotions during the study and test week were unrelated to both GPA and test score, but emotions during the feedback week were related to GPA and

test score. The only exception to this was anxiety where greater changeability in anxiety during the test week was related to reduced test performance.

4. 6 Significance of the Study

This study contributes to AE research by being the first to conduct an intensive longitudinal, repeated measure study around a whole assessment process allowing in-the-moment measurement of the students' emotions rather than relying on retrospective recall. While the use of repeated measure studies in AEs research exists, these are generally restricted to two or three time points. Although there has been a recent longitudinal AE study (Ahmed et al., 2013), it used fewer data collection points than the current study and was conducted over three school terms with a focus on the overall course achievement rather than on a particular assessment. Hence, the use of an intensive longitudinal study around one assessment event to capture emotion fluctuations is a distinguishing feature of the current study. To the best of our knowledge, the changeability of AEs and its implications on academic achievement has only been studied by Ahmed and colleagues (2013), and the relationship between the changeability of AEs to academic ability (i.e., GPA) has not been studied previously.

A further substantial contribution of this study is the testing and elimination of two conceptual approaches to understanding the structure of AEs. Our analysis rejects the power of valence (positive and negative emotions) and effect to fit the data and to relate to academic performance. Instead our study shows that AEs from three different sources have a conceptual and substantive meaning across the frameworks. We suggest that this approach has greater validity and power to help educational psychological research than the current conceptualisation. Nonetheless, we admit that this contribution depends heavily on chance artefacts in our data design and analysis and this claim requires further testing before it can be taken as but a more promising angle of attack into understanding student AEs.

4.7 Practical Implications

As the only emotion to relate to test score leading up to the test was anxiety, our findings suggest that a small amount of anxiety is an activator for greater student achievement. Students with low anxiety in relation to an upcoming test might benefit from (1) (2) a conversation about how the assessment is not easy, but it is not impossible to get a good mark. This might help increase the value of the course and the perceived demand of the assessment and lead to greater effort. Those with high levels of anxiety could also be reminded that anxiety, if controlled, can be an important motivator. Indeed, it has been demonstrated among New Zealand high school students that those with increased negative test anxiety 'tension' achieved a higher GPA (Chin, 2014). This result means that being tense, as opposed to worried, is not only normal, but potentially beneficial for better test performance.

4.8 Future Studies

For the purpose of LCM analysis, the sample size ($N = 166$) used in the current study was less than ideal and, hence, replication with larger samples is warranted. However, although diary study researchers have recommended a minimum of 30 participants, published diary studies tend to have samples of at least 100 participants (Ohly, Sonnentag, Niessen & Zapf, 2010).

The participants of this study were also recruited from a general education course, which is both an advantage and a limitation. The course included students representing a wide community of tertiary students which may enable generalisability of the study findings, but such generalisations may be threatened by insufficient sample within each discipline. Conducting a similar study with more students specialising in different subjects and with different assessment types with different contribution to course success may produce different results. Therefore, future research should investigate the universality of the current

study finding as well as exploring whether factors such as gender, ethnicity, study major, years at tertiary education, and age influence the results.

The current study also examined the changeability of emotions in three week blocks due to a lack of fit being found across the 9 day models for each emotion. To fully examine the changeability of the emotions across an assessment event, if sufficient model fit can be found, it would be better to create one model across the assessment period. However, our results may reflect the inherent reality that emotions are not linear across a complex process such as preparing for, undergoing, and then receiving results from an assessment.

There were two noteworthy methodological limitations in the current study. Due to the intensive nature of the design, the study may have attracted a particular type of participant. While the data screening sought to maximise the validity of any estimation of missing values, it did lead to the final sample having higher GPAs than the full sample. However, it is important to note that the current study did still contain the full range of GPAs (0-9) and the students had a wide range of test score 42- 92%. Further, as long as research depends on voluntary participation, it will be difficult to overcome this validity threat.

When constructing a psychological measurement, randomising the order of items and including alternative or negatively-worded items are recommended to minimise response style effects. The format of the current study's instrument listed the positive emotions first, followed by the negative emotions, and there were no reverse-worded items because research suggests very different psychological processes are invoked when trying to answer negatively worded items (Brown, 2004). If this instrument is to be used in future research, the order of positive and negative emotions should be randomised.

One thing that was missing from the instrument used in the study was space for the participants to explain why they were experiencing more or less of certain emotions at different time points and the value they placed on the upcoming assessment, the extent to which they perceived the upcoming assessment as demanding and how much control they had over their test performance. Adding room for contextual explanations for the emotions and the assessment may allow a deeper understanding of the relationship between appraisals (the CVT; Pekrun, 2006) and AEs in the different phases of assessment (e.g., before and after) and feedback.

4.9 Conclusion

This study provides a meaningful contribution to our understanding of AEs in an assessment context by conducting an intensive longitudinal diary method to study AEs, which had been lacking in the AE research field. The findings of the current study add to the CVT (Pekrun, 2006) by (1) identifying the specific associations between students' academic ability (GPA) and AEs, and (2) identifying of the association between students' AEs and actual test scores. In particular, the current study highlights that both GPA and Test score are most related to emotions in the feedback week once the results are known.

5.0 References

Ahmed, W., van der Werf, G., Kuyper, H., & Minnaert, A. (2013). Emotions, Self-Regulated Learning, and Achievement in Mathematics: A Growth Curve Analysis. *Journal of Educational Psychology, 105*(1), 150-161. doi: 10.1037/a0030160

Ainley, M. & Ainley, J. (2011). Student engagement with science in early adolescence: The contribution of enjoyment to students' continuing interest in learning about science. *Contemporary Educational Psychology, 36*, 4-12. doi: 10.1016/j.cedpsych.2010.08.001

Alexander, P. A., Murphy, P. K., Woods, B. S., Duhon, K. E., & Parker, D. (1997). College Instruction and Concomitant Changes in Students' Knowledge, Interest, and Strategy Use: A Study of Domain Learning. *Contemporary Educational Psychology, 22*(2), 125-146. doi: 10.1006/ceps.1997.0927

Aron, A., Aron, E. N., & Coups, E. (2005). *Statistics for Psychology* (4th ed.). New York: Prentice Hall.

Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 93-114). New York: Routledge.

Boekaerts, M. (1993). Anger in relation to school learning. *Learning and Instruction, 3*(4), 269–280. doi: 10.1016/0959-4752(93)90019-V

Bolger N., Davis A.,Rafaeli E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology , 54*, 579–616.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation approach.* Hoboken, NJ: Wiley.

Boomsma, A., & Hoogland, J. J. (2001). The robustness of LISREL modeling revisited. In R. Cudeck, S. Du Toit & D. Sorbom (Eds.), *Structural equation modeling: Present and future* (pp. 139-168). Lincolnwood, IL: Scientific Software International.

Brown, G. T. L. (2004). Measuring attitude with positively packed self-report ratings: Comparison of agreement and frequency scales. *Psychological Reports*, *94*, 1015-1024. doi: 10.2466/pr0.94.3.1015-1024

Buchtel, E. E. (2009). *A Sense of Obligation: Culture and the Subjective Experience of Meeting Expectations.* Unpublished doctoral dissertation, University of British Columbia, Vancouver, Canada.

Capa, R., & Audiffren, M. (2009). How does achievement motivation influence mental effort mobilization? Physiological evidence of deteriorative effects of negative affects on the level of engagement. *International Journal of Psychophysiology*, *74*(3), 236-242. doi: 10.1016/j.ijpsycho.2009.09.007

Chapell, M. S., Blanding, Z. B., Silverstein, M., Takahashi, M., Newman, B., Gubi, A., & McCann, N. (2005). Test Anxiety and Academic Performance in Undergraduate and Graduate Students. *Journal of Educational Psychology, 97*(2), 268-274. doi: 10.1037/0022-0663.97.2.268

Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods & Research, 29*(4), 468-508. doi: 10.1177/0049124101029004003

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255. doi: 10.1207/S15328007SEM0902_5

Chin, E. C.-H. (2014). *Emotional Determinants of Test Anxiety and Academic Performance.* (DClinPsych unpublished doctoral dissertation), Massey University, Palmerston North, NZ.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation, 10*(7), Available online: http://www.pareonline.net/pdf/v10n17.pdf.

Courtney, M. G. R. (2013). Determining the number of factors to retain in EFA: Using the SPSS R-Menu v2.0 to make more judicious estimations. *Practical Assessment Research & Evaluation, 18*(8), Available online: http://pareonline.net/getvn.asp?v=18&n=18.

Daniels, L. M., Haynes, T. L., Stupnisky, R. H., Perry, R. P., Newall, N., & Pekrun, R. (2008). Individual differences in achievement goals: A longitudinal study of cognitive, emotional, and achievement outcomes. *Contemporary Educational Psychology, 33*(4), 584–608. doi: 10.1016/j.cedpsych.2007.08.002

Daniels, L. M., Stupnisky, R. H., Pekrun, R., Haynes, T. L., Perry, R., & Newall, N. E. (2009). A Longitudinal Analysis of Achievement Goals: From Affective Antecedents to Emotional Effects and Achievement Outcomes. *Journal of Education Psychology, 101*(4), 948-963. doi: 10.1037/a0016096

DeLongis, A., Hemphill, K. J., & Lehman, D. R. (1992). A structured diary methodology for the study of daily events. In F. B. Bryant (Ed.), *Methodological Issues in Applied Social Psychology* (pp. 83-109). New York: Plenum Press

Deneen, C.C., Brown, G. T. L., Bond, T. G., & Shroff, R. H. (2013). Understanding outcome-based education changes in teacher education: Evaluation of a new instrument with preliminary findings. *Asia-Pacific Journal of Teacher Education*. doi: 10.1080/1359866x.2013.787392

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5*(3), 69-106.

Eysenck, M. W. (1979). Anxiety, learning, and memory: A reconceptualization. *Journal of research in personality*, *13*(4), 363-385.

Fan, Xitao, & Sivo, Stephen A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research, 42*(3), 509–529. doi: 10.1080/00273170701382864

Folkman, S., & Lazarus, R. S. (1985). If it changes it must be a process: A study of emotion and coping during three stages of a college examination. *Journal of Personality and Social Psychology, 48*(1), 150-170. doi: 10.1037/0022-3514.48.1.150

Goetz, T., Preckel, F., Pekrun, R., & Hall, N. C. (2007). Emotional experiences during test taking: Does cognitive ability make a difference? *Learning and Individual Differences, 17*, 3 – 16. doi: 10.1016/j.lindif.2006.12.002

Graham, S., Berninger, V., & Fan, W. (2007). The structural relationship between writing attitude and writing achievement in first and third grade students. *Contemporary Educational Psychology, 32*(3), 516-536. doi: 10.1016/j.cedpsych.2007.01.002

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research, 58*(1), 47–77. doi: 10.3102/00346543058001047

IBM. (2011). Amos [computer program] (Version version 20, Build 817). Meadville, PA: Amos Development Corporation.

Isen, A. M. (2000). Positive affect and decision making. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 417-435). New York: Guilford Press.

Isen, A. M., Daubman, K. A., & Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology, 52*(6), 1122–1131. doi: 10.1037/0022-3514.52.6.1122

Kitayama, S., Mesquita, B., & Karasawa, M. (2006). Cultural Affordances and Emotional Experience: Socially Engaging and Disengaging Emotions in Japan and the United States. *Journal of Personality and Social Psychology, 91*(5), 890-903. doi: 10.1037/0022-3514.91.5.890

Klockars, A. J., & Yamagishi, M. (1988). The influence of labels and positions in rating scales. *Journal of Educational Measurement, 25*(2), 85-96. doi: 10.1111/j.1745-3984.1988.tb00294.x

Koh, J., Shao, Y., & Wang, Q. (2009). Father, Mother and Me: Parental Value Orientations and Child Self-identity in Asian American Immigrants. *Sex Roles, 60,* 600-610. doi:10.1007/s11199-008-9550-z

Lam, T. C. M., & Klockars, A. J. (1982). Anchor point effects on the equivalence of questionnaire items. *Journal of Educational Measurement, 19*(4), 317-322. doi: 10.1111/j.1745-3984.1982.tb00137.x

Larson, R. W. (1989). Emotions and the creative process: Anxiety, boredom, and enjoyment in creative writing. Imagination, *Cognition and Personality, 9*(4), 275-292. doi: 10.2190/XT9G-WXRF-BK4M-36AK

Lichtenfeld, S., Pekrun, R., Stupnisky, R. H., Reiss, K., & Murayama, K. (2012). Measuring students' emotions in the early years: The Achievement Emotions Questionnaire-Elementary School (AEQ-ES). *Learning and Individual Differences, 22*(2), 190-201. doi: 10.1016/j.lindif.2011.04.009

Linnenbrink, E. A. (2007). The role of affect in student learning: A multi-dimensional approach to considering the interaction of affect, motivation, and engagement. In P. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 107-124). San Diego, CA: Academic Press. doi: 10.1016/B978-012372545-5/50008-3

Linnenbrink-Garcia, L., Rogat, T. K., & Koskey, K. L. K. (2011). Affect and engagement during small group instruction. *Contemporary Educational Psychology, 36*(1), 13-24. doi: 10.1016/j.cedpsych.2010.09.001

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Marsh, H W, Hau, K., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*(3), 320-341. doi: 10.1207/s15328007sem1103_2

McArdle, J. J. (2007). Five steps in the structural factor analysis of longitudinal data. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 99-130). Mahwah, NJ: LEA.

Meinhardt, J., & Pekrun, R. (2003). Attentional resource allocation to emotional events: An ERP study. *Cognition and Emotion, 17*(3), 477-500. doi: 10.1080/02699930244000039

Mouratidis, A. A., Vansteenkiste, M., Sideridis, G., & Lens, W. (2011). Vitality and Interest–Enjoyment as a Function of Class-to-Class Variation in Need-Supportive Teaching and Pupils' Autonomous Motivation. *Journal of Educational Psychology, 103*(2), 353-366. doi: 10.1037/a0022773

Nett, U. E., Goetz, T., & Hall, N. C. (2011). Coping with boredom in school: An experience sampling perspective. *Contemporary Educational Psychology, 36*(1), 49-59. doi: 10.1016/j.cedpsych.2010.10.003

Ohly, S., Sonnentag, S., Niessen, C., & Zapf, D. (2010). Diary studies in organizational research. *Journal of Personnel Psychology*, *9*(2), 79-93.

Owens, M., Stevenson, J., Hadwin, J. A., & Norgate, R. (2014). When does anxiety help or hinder cognitive test performance? The role of working memory capacity. *British Journal of Psychology, 105,* 92-101. doi: 10.1111/bjop.12009

Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review, 18*(4), 315-341. doi: 10.1007/s10648-006-9029-9

Pekrun, R., Elliot, A. J., & Maier, M. A. (2006). Achievement Goals and Discrete Achievement Emotions: A Theoretical Model and Prospective Test. *Journal of Educational Psychology, 98*(3), 583-597. doi: 10.1037/0022-0663.98.3.583

Pekrun, R., Elliot, A. J., & Maier, M. A. (2009). Achievement Goals and Achievement Emotions: Testing a Model of Their Joint Relations With Academic Performance. *Journal of Educational Psychology, 101* (1), 115-135. doi: 10.1037/a0013383

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). Chapter 2 - The Control-Value Theory of Achievement Emotions: An Integrative Approach to Emotions in Education. In A. S. Paul & P. Reinhard (Eds.), *Emotion in Education* (pp. 13-36). Burlington: Academic Press.

Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., & Perry, R. P. (2010). Boredom in Achievement Settings: Exploring Control-Value Antecedents and Performance Outcomes of a Neglected Emotion. *Journal of Educational Psychology, 102*(3), 531-549. doi: 10.1037/a0019243

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology, 36*(1), 36-48. doi: 10.1016/j.cedpsych.2010.10.002

Pekrun, R., Goetz, T., Perry, R. P., Kramer, K., Hochstadt, M., & Molfenter, S. (2004). Beyond Test Anxiety: Development and Validation of the Test Emotions Questionnaire (TEQ). *Anxiety, Stress, and Coping, 17*(3), 287-316. doi: 10.1080/10615800412331303847

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37*(2), 91–106. doi: 10.1207/S15326985EP3702_4

Peterson, E. R., Brown, G. T. L., & Hamilton, R. J. (2013). Cultural differences in Tertiary Students' Conceptions of Learning as a Duty and Student Achievement. *The International Journal of Quantitative Research in Education, 1* (2), 167-181. doi:10.1504/IJQRE.2013.056462

Peterson, R. A. (1994). A Meta-Analysis of Cronbach's Coefficient Alpha. *Journal of Consumer Research, 21*(2), 381-391. doi:10.2307/2489828

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353.

Schutz, P. A., & Davis, H. A. (2000). Emotions and self-regulating during test taking. *Educational Psychologist, 35*(4), 243−256.

Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research, 4*(1), 27–41. doi: 10.1080/08917779108248762

Smith, C. A., & Ellsworth, P. C. (1987). Patterns of appraisal and emotion related to taking an exam. *Journal of Personality and Social Psychology , 52*(3), 475-488. doi: 10.1037/0022-3514.52.3.475

Turner, J. E., & Schallert, D. (2001). Expectancy-value relationships of shame reactions and shame resiliency. *Journal of Educational Psychology, 93,* 320–329. doi: 10.1037/0022-0663.93.2.320

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-70. doi: 10.1177/109442810031002

Weiner, B. (1985). An Attributional Theory of Achievement Motivation and Emotion. *Psychological Review, 92*(4), 548-573. doi: 10.1037/0033-295X.92.4.548

Wolters, C. A. (2003). Regulation of Motivation: Evaluating an Underemphasized Aspect of Self-Regulated Learning. *Educational Psychologist, 38*(4), 189-205. doi: 10.1207/S15326985EP3804_1

Zeidner, M. (2007). Chapter 10 – Test Anxiety in Educational Contexts: Concepts, Findings, and Future Directions. In A. S. Paul & P. Reinhard (Eds.), *Emotion in Education* (pp. 165-184). Burlington: Academic Press.

Zoller, U., & Ben-Chaim, D. (1988). Interaction between examination-type anxiety state, and academic achievement in college science: An action-oriented research. *Journal of Research in Science Teaching, 26*, 65-77.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores.* Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defence.

Table 1.
Valence and Effect of Emotions Extracted from the AEQ, Kitayama et al. (2006) and Buchtel (2009)

Valence & Effect	The AEQ	Kitayama et al. (2006)	Buchtel (2009)
Positive AEs			
Activating/ Engaging	Enjoyment (A)		
	Hope (A)		
	Pride (A)		
		Close feelings (E)	Close (E)
			Appreciated (E)
		Friendly feelings (E)	Friendly feelings (E)
		Respect (E)	Respect (E)
		Sympathy (E)	
Neutral / no direction specified		Calmness	Calm (N)
			Competent (N)
		Elation	Elated (N)
		Happy	Happy (N)
		Relaxation	Relaxed (N)
Deactivating/ Disengaging	Relief (De)		
		Pride (D)	Proud (D)
		Respected (D)	Self-respect (D)
		Superior (D)	Superior (D)
		Top of the world (D)	
Negative AEs			
Activating/ Engaging	Anger (A)		
	Anxiety (A)		
	Shame (A)	Ashamed (E)	Ashamed (E)
		Fear (E)	
		Guilt (E)	Guilty (E)
		Indebted (E)	
Neutral / no direction specified		Boredom	Bored (N)
		Depression	
		Disgust	Disgusted (N)
		Unhappy	Unhappy (N)
		Sadness	
			Annoyed (N)
Deactivating/ Disengaging		Frustration (D)	Frustrated (D)
	Hopelessness (De)		
		Sulky feelings (D)	Sulky feelings (D)
		Angry (D)	Angry (D)
	Boredom (De)		

Note. (A) indicates activating emotions, (De) indicates deactivating emotions, (E) indicates engaging emotions, (D) indicates disengaging emotions and (N) indicates neutral emotions classified by Kitayama et al. (2006) and Buchtel (2009).

Table 2.
Number of participants and proportion of final sample according to demographic characteristics

Characteristic	*N*	% of sample
Sex		
Female	119	71.70
Male	47	28.30
Age (Years)		
16-19	80	48.20
20-29	79	47.50
30+	5	3.00
Missing	2	1.20
Grade Point Average (GPA)		
A grades	30	18.07
B grades	89	53.61
C grades or lower	47	28.31
Academic Program		
Science	74	44.58
Business	46	27.71
Arts	20	12.05
Conjoint	15	9.04
Other	8	4.82
Engineering	2	1.20
Law	1	0.60
Ethnicity		
Asian	71	42.80
NZ European	60	36.10
Other	26	15.70
Pasifika	5	3.00
Maori (indigenous people of New Zealand)	2	1.20
Middle Eastern	2	1.20

Note. $N = 166$; this is *N* after participants with missing data were removed.

Table 3. Fit of the emotion models proposed by Pekrun, Kigtyama and Buchtel in both the 9 day administration model and the three week block (study, test and feedback week) model

Model	Description	Stats	Commentary
		Nine day administration models	
1a. Pekrun's AEQ	MGCFA invariance 9 days (*k*=81)	χ^2=983.80; *df*=386; χ^2/*df*=2.55; RMSEA=.032; CFI=.896; SRMR=.064; gamma hat=.96	Strong invariance; metric & scalar
1b. Kitayama	MGCFA invariance 9 days (*k*=207)	χ^2=6025.89; *df*=2229; χ^2/*df*=2.70; RMSEA=.034; CFI=.841; SRMR=.082; gamma hat=.98	Weak invariance; metric only (one error variance corrected to .005)
1c. Buchtel	MGCFA invariance 9 days (*k*=189)	χ^2=5513.12; *df*=1844; χ^2/*df*=3.13; RMSEA=.037; CFI=.839; SRMR=.079; gamma hat=.97	Weak invariance; metric only
2. All positive-negative	MGCFA invariance 9 days (*k*=279)	χ^2=11044.71; *df*=3897; χ^2/*df*=2.76; RMSEA=.032; CFI=.805; SRMR=.080; gamma hat=.97	Weak invariance; metric only
3a. Kitayama	Engage-disengage-neutral, MGCFA invariance 9 days (*k*=207)	χ^2=5675.90; *df*=2178; χ^2/*df*=2.61; RMSEA=.034; CFI=.854; SRMR=.073; gamma hat=.98	Weak invariance; metric only
3b. Buchtel	Engage-disengage-neutral, MGCFA invariance 9 days (*k*=189)	Negative error variance Negative Disengage CR=2.432	Inadmissible
		Three week models (study, test and feedback week)	
4a. AEQ Model 1a only	LCM Week 1 + GPA & test	positive-negative correlation covariance not positive definite in Day 3	Inadmissible
4b. Kitayama Model 1b only	LCM Week 1 + GPA & test	positive-negative correlation covariance not positive definite in Day 1 and Day 3	Inadmissible
4.c Buchtel Model 1c only	LCM Week 1 + GPA & test	positive-negative correlation covariance not positive definite in Day 1 and Day 3	Inadmissible
4d. All Model 2	LCM Week 1 + GPA & test (*k*=98)	χ^2=10324.65; *df*=4650; χ^2/*df*=2.22; RMSEA=.086; CFI=.610; SRMR=.191; gamma hat=.59	Reject fit; no statistically significant paths to GPA or Test score
4.e All Model 2	LCM Week 2 + GPA & test (*k*=98)	χ^2=11101.49; *df*=4650; χ^2/*df*=2.39; RMSEA=.092; CFI=.629; SRMR=.193; gamma hat=.55	Reject fit; no statistically significant paths to GPA or Test score
4f. All Model 2	LCM Week 3 + GPA & test (*k*=98)	χ^2=12699.35; *df*=4650; χ^2/*df*=2.73; RMSEA=.102; CFI=.602; SRMR=.266;	Reject fit; no statistically significant paths to GPA or Test

		gamma hat=.50	score

Note. MGCFA=multi-group confirmatory factor analysis; *k*=number of manifest items in model; RMSEA=root mean square error of approximation; CFI=comparative fit index; SRMR=standardized root mean residual; GPA=grade point average; LCM=latent curve model.

Table 4.

Factor Pattern Matrix Showing the Factor Loadings for Positive and Negative Emotions

	Factor Pattern Matrix			Factor Pattern Matrix		
Positive Emotion Items	Happy	Chilled	Negative Emotion Items	Sad	Anxious	Self-loathing
I felt happy.*#	**.96**	-.07	I felt sad. *#	**.87**	-.11	-.09
I experienced enjoyment.* #	**.92**	-.12	I was unhappy. *#	**.84**	-.04	-.08
I felt friendly feelings.* #	**.90**	-.08	I felt depressed. *#	**.75**	.07	-.04
I felt appreciated.* #	**.69**	.08	I felt sulky feelings.* #	**.71**	.16	-.01
I felt like I was respected.* #	**.59**	.20	I felt like I was indebted.	.33	.26	-.07
I felt competent.*	**.43**	.25	I felt frustrated.* #	.13	**.71**	-.03
I felt elated.	.54	.33	I felt anxious.* #	.30	**.55**	.07
I felt hopeful.	.50	.36	I was fearful.* #	.19	**.47**	-.22
I felt close feelings.	.40	.21	I was bored.*	-.07	**.41**	-.02
I felt superior.*	-.16	**.90**	I felt annoyed.	.03	.47	-.35
I felt relieved.*	.05	**.77**#	I felt ashamed.* #	-.09	.03	**-.87**
I felt like I was on the top of the world.*	.30	**.50**	I felt disgusted.* #	.21	-.16	**-.74**
I felt relaxed.*	.29	**.48**#	I felt angry.* #	.04	.06	**-.64**
I felt sympathetic.*	.03	**.47**	I felt guilty.* #	.11	.17	**-.51**
I felt calm.*	.21	**.46**#	I felt hopeless.	.31	.23	-.38
I felt proud.	.43	.46				
I felt self-respect.	.39	.44				
Factor inter-correlations			Factor inter-correlations			
Happy		.70	Sad		.50	-.69
			Anxious			-.53

Notes. * indicates items that were kept in the initial factor analysis and used in the initial CFA. # = items that remained after trimming low loading items in the CFA to improve model fit.

Table 5

Multi-group invariance testing per emotion factor across nine administration days.

	CFI				ΔCFI		
Emotion	Unconstrained	Metric	Scalar	Residual	Metric	Scalar	Residual
Positive							
Chilled	1.000	0.998	0.893	0.891	**0.002**	0.105	0.002
Happy	0.972	0.972	0.961	0.961	**0.000**	0.011	0.000
Negative							
Anxious	1.000	0.999	0.912	0.909	**0.001**	0.087	0.003
Self-loathing	0.989	0.975	0.959	0.955	0.014	0.016	0.004
Sad	0.998	0.989	0.980	0.979	**0.009**	**0.009**	**0.001**

Note. CFI = comparative fit index; ΔCFI<.01 shown in bold.

Table 6.

Fit Statistics for LCM Models for Five Emotions Across the Nine Days and the Emotions in Three Week Blocks.

	Fit Statistics						
Emotion & Model	χ^2	*df*	χ^2/*df* (*p*)	CFI	RMSEA	SRMR	Gamma hat
Happy (9 day model)	*2010.90*	*918*	*2.19 (.14)*	*.856*	*.085*	*.076*	*.77*
Study Week	190.93	77	2.48 (.12)	.942	.095	.062	.92
Test Week	220.63	78	2.83 (.09)	.929	.105	.065	.90
Feedback Week	185.51	77	2.41 (.12)	.961	.092	.048	.92
Chilled (9 day model)	*573.66*	*297*	*1.93 (.16)*	*.910*	*.075*	*.086*	*.90*
Study Week	46.96	18	2.61 (.11)	.960	.099	.060	.96
Test Week	43.14	19	2.27 (.13)	.970	.088	.058	.97
Feedback Week	71.77	18	3.99 (.05)	.949	.135	.052	.93
Anxious (9 day model)	*727.85*	*297*	*2.45 (.12)*	*.86*	*.094*	*.076*	*.84*
Study Week	18.59	18	1.03 (.31)	.999	.014	.030	.99
Test Week	67.32	20	3.37 (.07)	.935	.120	.069	.94
Feedback Week	114.97	20	5.75 (.02)	.914	.170	.056	.89
Sad(9 day model)	1145.71	*559*	*2.05 (.15)*	*.91*	*.080*	*.080*	*.84*

	Fit Statistics						
Emotion & Model	χ^2	*df*	χ^2/df *(p)*	CFI	RMSEA	SRMR	Gamma hat
Study Week	99.46	43	2.31 (.13)	.964	.089	.049	.95
Test Week	85.84	45	1.91 (.17)	.979	.074	.072	.96
Feedback Week	154.40	43	3.59 (.06)	.952	.125	.055	.98
Self-loathing (9 day model)	*1104.10*	*559*	*1.98 (.16)*	*.87*	*.077*	*.075*	*.85*
Study Week	97.41	44	2.21 (.14)	.944	.086	.059	.95
Test Week	83.41	43	1.94 (.16)	.961	.075	.045	.96
Feedback Week	169.41	44	3.85 (.05)	.919	.131	.055	.89

Note. 9-day fit values shown in italics.

Table 7
LCM Intercept and Slope as Predictors of the Dependent Test Score for each Emotion Factor by Assessment Week

	Week in Assessment Process					
	Study Week		Test Week		Feedback Week	
Emotion	Intercept	Slope	Intercept	Slope	Intercept	Slope
Positive						
Happy	0.06	0.23	0.12	0.24	0.26 **	-0.14
Chilled	0.05	0.28	0.07	0.19	0.34 *	-0.24
Negative						
Anxious	0.10	-0.11	0.10	-0.46	—	—
Sad	-0.11	-0.25	0.03	0.05	-0.21*	-0.05
Self-loathing	-0.11	-0.11	-0.31	-0.35	-0.24**	0.01

Note. All values are standardised regression weights β; no values calculated for Anxious Feedback week because of poor model fit; ** = p<.01; *=*p*<.05.

Table 8
Regression Weight of GPA on the Intercept and Slope for each emotion factor

	Week in Assessment Process					
	Study Week		Test Week		Feedback Week	
Emotion	Intercept	Slope	Intercept	Slope	Intercept	Slope
Positive						
Happy	-0.02	0.04	0.06	0.14	0.21*	-0.14
Chilled	-0.02	0.09	0.01	0.02	0.22*	-0.21
Negative						
Anxious	0.05	-0.13	0.15	-0.48	—	—
Sad	-0.01	-0.07	-0.02	0.15	-- 0.18 *	0.06
Self-loathing	-0.08	0.02	-0.09	-0.05	-0.20*	-0.03

Note. All values are standardised regression weights β; no values calculated for Anxious Feedback week because of poor model fit; *=p<.05

Table 9

GPA as a predictor of Test score and the Slopes and Intercept, and the slopes and intercepts as predictors of test score in the Feedback Week

	Standardised Regression Weights and Effects						Fit Statistics						
Feedback Week Emotions	GPA on Test (β)	GPA on Intercept (β)	GPA on Slope (β)	Intercept on Test (β)	Slope on Test (β)	Test SMC	χ^2	*df*	χ^2/df *(p)*	CFI	RMSEA	SRMR	Gamma hat
Positive Emotion													
Happy	.54***	.17*	ns	.18 *	ns	.36	211.92	105	2.02(.16)	.96	.08	.046	.93
Chilled	.55***	ns	ns	.20**	ns	.35	92.114	25	2.63(.11)	.95	.09	.064	.96
Negative Emotion													
Sad	.57***	-.15*	ns	ns	ns	.33	169.57	66	2.57(.11)	.96	.10	.052	.91
Self-loathing	.57***	-.21**	ns	ns	ns	.33	187.67	67	2.80(.09)	.93	.10	.055	.91

Note. All values are standardised regression weights β; no values calculated for Anxious Feedback week because of poor model fit; SMC = squared multiple correlation or R^2; ** = p<.01; *=*p*<.05, ns=not significant

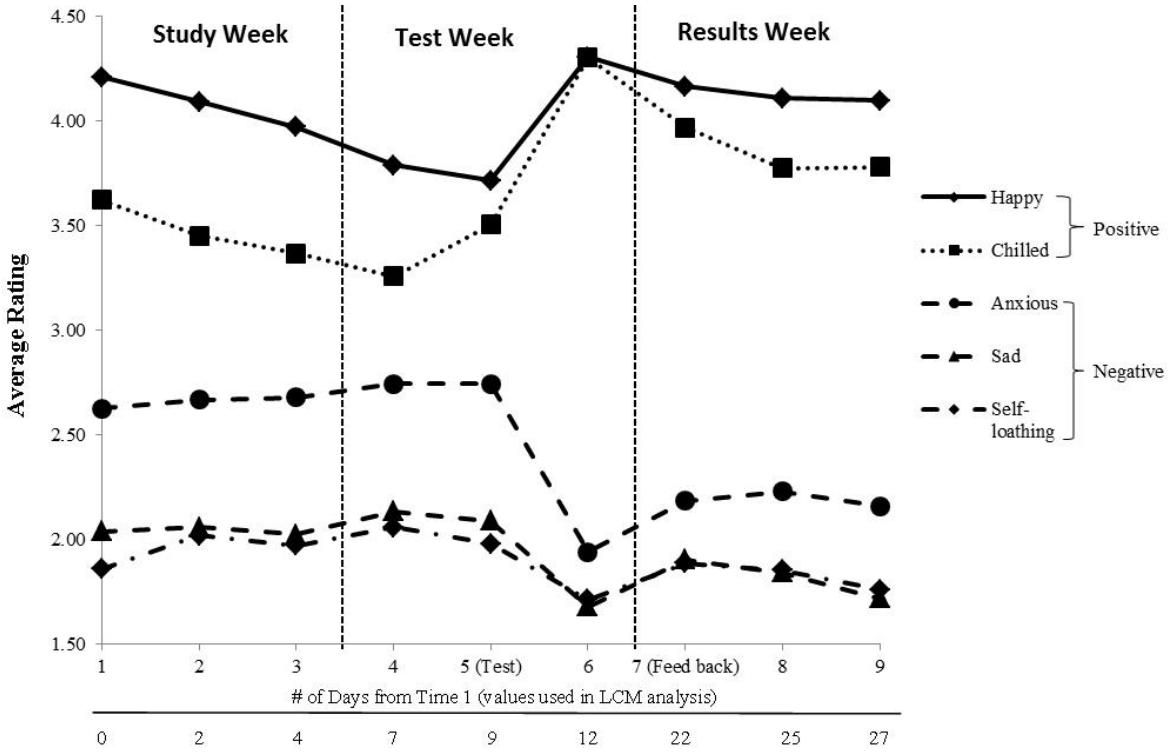

Figure 1. Mean Factor Scores of the Five Emotions over Nine Administration Days.

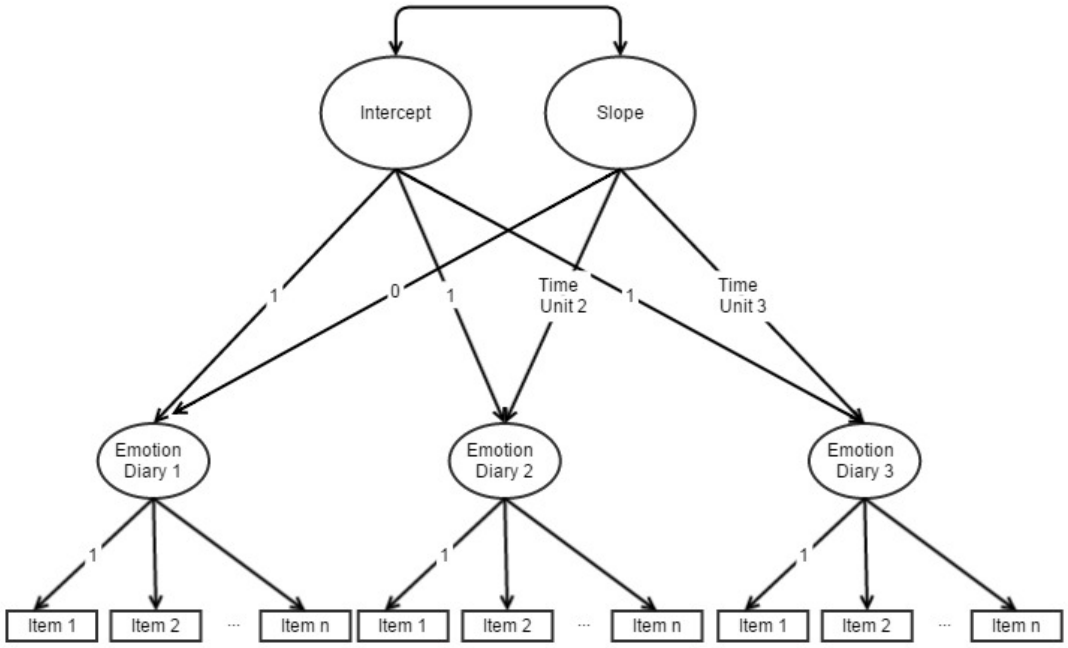

Figure 2. A schematic model of a week-long latent curve model used for each emotion factor in the study

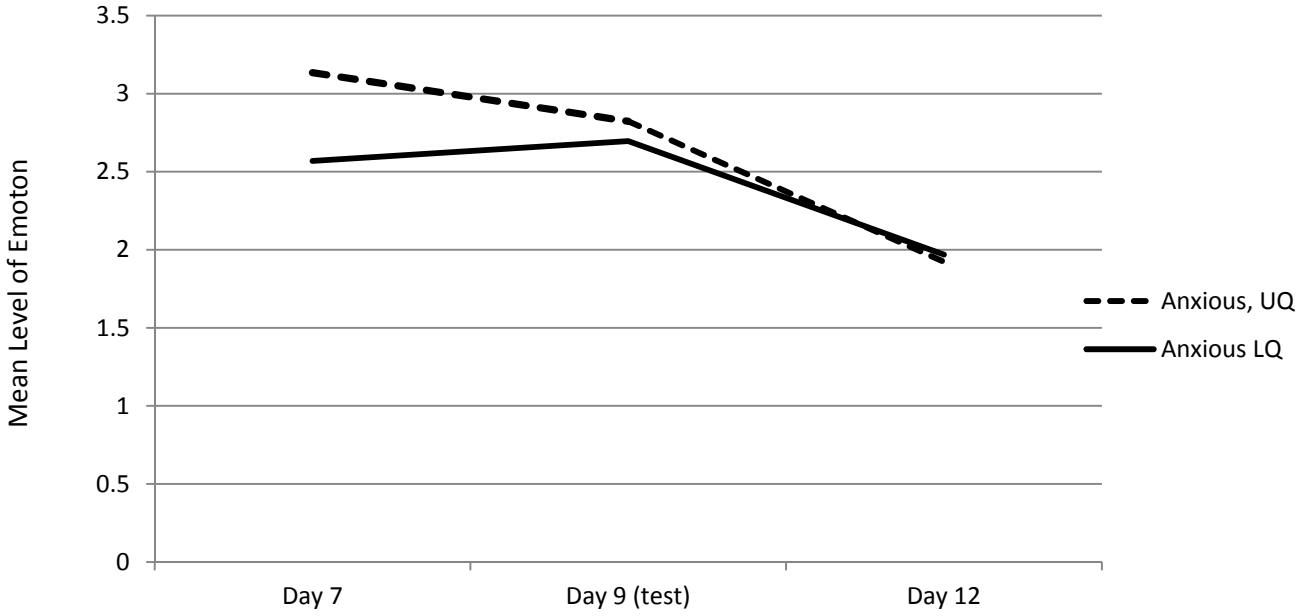

Figure 3. Mean levels of anxiety for students performing on upper quartile (UQ) and lowest quartile (LQ) on the test

Supplementary Appendix Table 1.
Fit Statistics for LCM Models predicting test score for each emotion factor on each Week (study, test, feedback) of the Assessment

	Fit Statistics						
Emotion & Model	χ^2	*df*	χ^2/df (*p*)	CFI	RMSEA	SRMR	Gamma hat
Happy							
Study Week	210.04	90	2.33 (.13)	.94	.09	.060	.92
Test Week	230.72	91	2.63 (.10)	.93	.10	.063	.90
Feedback Week	201.21	90	2.24 (.13)	.96	.09	.046	.92
Chilled							
Study Week	54.22	25	2.17(.14)	.96	.08	.058	.97
Test Week	62.84	26	2.42(.12)	.96	.09	.065	.96
Feedback Week	83.94	25	2.26 (.13)	.94	.12	.049	.93
Anxious							
Study Week	25.80	25	1.03 (.31)	1.00	.01	.035	1.00
Test Week	81.33	26	3.01(.08)	.93	.11	.069	.94
Feedback Week	-	-	-	-	-	-	-
Sad							
Study Week	110.47	53	2.08(.14)	.96	.09	.049	.95
Test Week	101.21	55	1.84(.18)	.98	.07	.074	.96
Feedback Week	263.59	52	3.09(.08)	.95	.11	.053	.91
Self-loathing							
Study Week	100.65	54	1.86(.17)	.95	.07	.058	.96
Test Week	93.59	53	1.77(.18)	.96	.07	.043	.96
Feedback Week	177.10	54	3.28(.07)	.92	.12	.052	.89

Supplementary Appendix Table 2.
Fit Statistics for LCM Models Examining the Regression Weight of GPA on the Intercept and Slope for each emotion factor

	Fit Statistics						
Emotion & Model	χ^2	*df*	χ^2/df (*p*)	CFI	RMSEA	SRMR	Gamma hat
Happy							
Study Week	213.302	90	2.37 (.12)	.938	.091	.062	.91
Test Week	240.269	91	2.64 (.10)	.926	.100	.063	.96
Feedback Week	194.072	90	2.16 (.14)	.962	.084	.046	.93
Chilled							
Study Week	56.260	25	2.25 (.13)	.957	.087	.059	.97
Test Week	53.812	26	2.07 (.15)	.966	.081	.058	.97
Feedback Week	77.310	25	3.09 (.08)	.950	.113	.049	.95
Anxious							
Study Week	27.465	25	1.10 (.29)	.996	.024	.037	1.00
Test Week	87.056	26	3.35 (.07)	.917	.119	.070	.93
Feedback Week							
Sad							
Study Week	112.462	53	2.12 (.15)	.962	.082	.050	.95
Test Week	94.107	55	1.71 (.19)	.980	.066	.066	.96
Feedback Week	157.958	53	2.98 (.08)	.954	.110	.051	.91
Self-loathing							
Study Week	105.577	54	1.96 (.16)	.946	.076	.057	.95
Test Week	96.750	53	1.83 (.18)	.958	.071	.046	.96
Feedback Week	174.154	54	3.23 (.07)	.923	.116	.052	.90