Introducing block mode to first-year university students: a natural experiment on satisfaction and performance

Introducing block mode to first-year University students: A natural experiment on satisfaction and performance

Daniel Loton[1], Cameron Stein[2], Philip Parker[3] & Mary Weaven[4]

[1]Connected Learning, Victoria University, Melbourne, Australia

[2]Data Insights, Victoria University, Melbourne, Australia

[3]Institute for Positive Psychology and Education, Australian Catholic University

[4]College of Education, Victoria University, Melbourne, Australia

Author Note

Daniel Loton https://orcid.org/0000-0003-4106-0555

Philip Parker https://orcid.org/0000-0002-4604-8566

We thank the Connected Learning and Data Insights Departments for supporting the research, Ms. Leisa Franklin for extracting assessment design data, the authors of all R packages utilised, and individuals consulted during manuscript preparation, specifically: Profs. Ian Solominides and Andrew Smallridge, Mx. Trish McCluskey, and A/Prof. Maxwell Winchester.

Statement of contributions: DL conceived the study, obtained ethical approval, undertook the literature review, analysis, and drafted all sections of the manuscript; CS navigated and helped interpret institutional data, undertook the SES and QILT analysis, and reviewed the draft manuscript; PP guided all aspects of the statistical analyses and use of the specific software applications utilised, and reviewed the draft manuscript; and MW expanded the literature review.

Three authors are employed by the intervention institution.

Correspondence concerning this article should be sent to Daniel.Loton@vu.edu.au.

Abstract

A multidisciplinary Australian University introduced a block model of blended, sequential four-week blocks to first-year students. This natural experiment compares the inaugural block and two prior cohorts on satisfaction and performance (n = 15,989 satisfaction and n = 86,545 assessment observations). Mixed effect cross-classified models with comprehensive controls, moderation and sensitivity testing show substantial increases in performance, especially for equity groups important in an expanding sector: low-socioeconomic and prior academic achievement, non-English speaking background. Effects on satisfaction were small with increases in teaching satisfaction but decrement in course satisfaction, especially perceived reasonable workload. Discipline consistently moderated effects and units that redesigned assessments offset some decline in course satisfaction. New models of higher education may support improved outcomes.

Keywords: Block mode, intensives, student satisfaction, assessment outcomes, equity indicators

Introducing block mode to first-year university students: A natural experiment on satisfaction and performance.

"This remissioning of the university...[is] challenging, and changing, the temporalities of university life" (Vostal & Roberston, 2012, p. 5).

An enduring aspect of university education is how study is structured, often termed the mode or model. Students typically undertake multiple units, also termed modules or courses, concurrently over the length of a semester in a combination of lectures and tutorials (terminology varies). Like other aspects of the university, study mode is also changing. Notable shifts include intensives (Scott & Conrad, 1992), blended and online learning (Vo et al., 2017), and active learning which can involve reducing or replacing lectures (Dawson & Dawson, 2018). Intensives, where students undertake a semester-length unit of study in a much shorter time, but without competing demands from parallel units, are most relevant to the present study (Wlodkowski, 2003).

Similar to intensives but offered as the primary format is block, where entire qualifications are delivered in sequential, short duration units. Block mode is rare in Western higher education, mostly adopted in small-scale liberal arts colleges. Studies of introducing block mode at these colleges (Heist & Taylor, 1979; Vaughan & Carlson, 1992) lacked pre-intervention data, hence the effects block mode are not well understood, especially when scaled to larger student cohorts.

In 2018 a university of over 28,000 students began replacing semester-length lectures and tutorials with successive four-week blocks: blended workshops studied sequentially. This new arrangement was delivered to first-year Bachelor students in 2018. We treat introduction of this

new model as a natural experiment and investigate effects on student satisfaction and performance (academic achievement), by comparing the inaugural first-year cohort with counterparts from two prior years. We employ multilevel cross-classified models suitable to structures in the data, and that control for differences across student cohorts and the effect of different teachers. Sensitivity tests probe the consistency of results across different sampling and analysis choices, and moderator models test whether any student sub-groups are 'left behind' by introduction the block.

The Times They Are A Changin': New Higher Education Models

Extensive scholarship considers the changing nature of Western university education. Increasing access and size, instrumental application of social planning and economic growth, privatisation, declining government funding and growing student costs, increasing staff casualisation and managerialism are contemporary themes (Barnett, 2019; Marginson & Considine, 2000; Trow & Burrage, 2010). Universities are sometimes derided as idiosyncratic and insular, inefficient, lacking in innovation, and staid (Christensen & Eyring, 2011; Helfand, 2013, 2016; Manicas, 2007). Others highlight how Universities, in particular through research, can be seen as "the birthplace of many new technologies" (Walker, 2009, p. 491), and engine of economic, social and cultural growth (Dawkins, 1987). One of the most enduring and distinctive aspects of university is the mode of study, but this is now also changing.

Expansion is probably the most important trend in the university sector, termed 'massification' in the US (Gumport et al., 1997) or the transition from elite to mass in Australia (Bradley et al., 2008), dating back to at least 1987 (Organisation for Economic Co-operation and Development [OECD], 1987; Dawkins, 1987). In fact Australia currently has the highest tertiary study admittance rate of all OECD nations at 77% in 2016 (OECD, 2019). Expansion entails

servicing a growing number of students sometimes described as ‘non-traditional’ (Baik et al., 2015) or the ‘new majority’ (Keller, 2001), who may not be as prepared for certain university experiences as a more privileged historical minority. Demography and automation-driven industrial revolution is likely to increase the number of people seeking further and higher education (Gleason, 2018; OECD, 2008).

Universities globally are adjusting to cater for this new majority. Changes include alternative entry criteria (van Ooijen-van der Linden et al., 2018), extra-curricular support programs to enhance engagement and retention (Harackiewicz & Priniski, 2018), often targeted with increasing sophistication (Vytasek et al., 2020), and extensive efforts to monitor and improve teaching and learning quality. The latter includes internal quality assurance, regulatory oversight, and national and international rankings.

Australia publishes rankings based on an annual survey of university student satisfaction, in hopes of driving quality through consumer choice (QILT, 2017). Some improvements are evident. A non-matched panel of over two decades suggests overall improvement in experiences of first-year students transitioning from secondary school to university (Baik et al., 2015). However in targeted samples from the same study, students with a lower academic entry score remain at higher risk of attrition, are less prepared, engaged and more stressed (Baik et al., 2019; Naylor et al., 2018).

Universities are also adopting distinctive models of teaching and learning to cater for this new majority. Intensives, block, blended or flipped classroom, active learning and seminars are examples of ‘models’ (Davies, 2006), ‘formats’ (Lutes & Davies, 2018), ‘modes’ (Male et al., 2016), ‘interventions’ or more simply “particular learning and teaching approaches” (Dawson & Dawson, 2018, p. 1405). Such models are characterised by a distinctive design and delivery of

curricula, with accompanying teaching approaches and activities; and are increasingly a focus of research (Tight, 2019). In this paper, we characterise block mode as a university educational intervention aimed at improving student success and satisfaction (Hsieh et al., 2005).

Blended and Active Learning

Intensives are the most directly comparable model to block mode. However blended and active learning, which prioritise small-group interaction and can reassign lecture time to pre and post-class online content (Fulton, 2012; Vo et al., 2017), contain similar elements. In active learning (Crouch & Mazur, 2001), for which many alternative terms are used (see Froyd & Simpson, 2008), change may occur primarily within the existing structure, for example by introducing audience response systems or interactivity windows, but otherwise retaining the duration of lectures (Gauci et al., 2009; Huxham, 2005). Other active learning interventions reduce lecture time in favour of small-group interactive classrooms (Cobain & Newberry, 2009), and some remove lectures altogether (King et al., 2018).

At least three meta-analyses have concluded blended learning significantly and moderately outperforms more traditional university models on academic achievement, especially for the STEM disciplines, and when online environments were used for cognitive rather than content support (Bernard et al., 2014; Vo et al., 2017). The result was robust across different assessment task types (Means et al., 2014). After correcting for publication bias active learning had a small significant meta-effect on student performance (Dawson & Dawson, 2018). Fidelity or consistency in the definition, design and implementation of educational models vary, and aspects of models often overlap (see, Aditomo et al., 2013). Meta-analyses have not yet investigated intensives or block mode.

Intensives

Intensives, or synonymously accelerated, time-compressed, short-cycle, shortened, ‘immersion scheduling’ and summer/winter units are growing in prevalence. Proponents list many potential benefits such as increased focused learning, control over scheduling with fewer conflicting demands, in-depth discussion, memorable experiences, greater assessment performance, enhanced peer-peer and teacher-peer relations; with decreased ‘extraneous material’ and procrastination (Jansen, 2004; Male et al., 2016; Scott, 2003; Wlodkowski, 2003). Potential risks include reduced course rigour, exhaustion for staff and students, more rapid falling behind with absence as “missing a day of class is like missing almost a week on the semester system” (Vaughan & Carlson, 1992, p. 266), and lack of timely feedback (Male et al., 2016; Wlodkoswki, 2003; Lutes & Davis, 2013).

Per a now apparently defunct Centre for the Study of Accelerated Learning, there were at least 320 intensive courses in the United States and Canada in 2006 (David, 2006). Some reviews state intensives are most frequently employed in the UK and Canada, and in particular in the disciplines of business and economics (Daniel, 2000; Davis, 2006; Marques, 2012), a point supported by a recent audit of Australian university offerings (Harvey et al., 2017). In a survey of Australian-based course coordinators Male et al. (2016) found 52% of intensives were taught at both undergraduate and postgraduate levels and offered to fit study with outside activities (30%), promote interactive learning (25%), and so students could focus on one unit at a time (10%).

Studies on intensives date back to at least 1960 (Scott & Conrad, 1992), and comprise empirical comparative studies, case studies, and theoretical works of pedagogical / andragogical

and curriculum considerations (Scott, 2003; Wlodkowski, 2006; Marques, 2012). Scott (2003) identified four pivotal categories for successful intensive design: characteristics of instructors, teaching approaches, classroom environment and assessment. To summarise, Scott stresses the use of active/interactive, experiential and applied learning, depth over breadth in content, close teacher-student and student-student relationships, and altered assessment design suitable to shorter timeframe, constructively aligned with few exams.

Several comparative studies have investigated intensives with a focus on student performance. Summarising an earlier review (Scott & Conrad, 1992), Davis (2006) reported most studies found no difference in performance. Davis also noted common research limitations including student self-selection into intensives, with no controls or random assignment employed despite systematic differences in students choosing to study intensive mode (Burton & Nesbit, 2008), and a lack of assessed longitudinal outcomes such as employment. Subsequent to 1992 some large-sample and more rigorously designed studies on intensives have been undertaken.

In a survey of n=1089 postgraduate students studying accounting, human resourcing and marketing (Burton & Nesbit, 2008), 48.8% to 26.3% chose to study in intensive mode, with more experienced and higher subject load students studying HRM the most likely to choose intensives. When asked why students mainly reported lifestyle factors such as accommodating work and travel, perception of the subject being easier or more familiar. Smaller numbers (18.7%) stated some educational benefit of traditional mode, such as more time to absorb material, and lesser number (6.7%) stated educational benefits of intensives. The authors defined traditional mode (termed ‘weekly’) somewhat differently from semester-length subjects.

Lutes and Davies (2013) analysed institutional student data at Brigham Young University. Using an initial sample of 29,000 records, students in intensives spent significantly less total time studying (time use is central to the US credit system), but by an amount deemed negligible at just over an hour more per week in semester mode. In a subsample of cases with matched instructors they found a smaller effect size of just under an hour. Extending this work Lutes and Davies (2018) found the level of teacher autonomy moderated the value students placed on homework. A small-sample (n=36) survey of lecturers indicated most units did not change greatly in syllabi across mode, that some perceived reading and writing-heavy units were less suited to intensives, and most reported increased efficiency but potentially limited deep learning.

Austin and Gustafson (2006) analysed institutional data from the University of West Georgia comprising academic achievement records of 11,795 students from 2001 to 2004, comparing semester units to intensives of varied duration. After controlling for student demographic, educational readiness and study choice factors, and undertaking some sensitivity testing, results were robust in suggesting academic achievement was greatest at four-week duration. As a novel test of whether improved results reflected lower assessment standards, the authors examined subsequent academic performance. In subsamples of directly linked units in the topics of accounting, math, foreign language and economics, and controlling for intervening units, intensive format did not moderate the effect of prior grades on subsequent grades, suggesting no reduction in learning or assessment standards.

In a rare study of student satisfaction, Kucsera and Zimmaro (2010) compared results on the Course Instructor Survey across intensive (nine week and 11 week) and traditional formats (15 weeks), for five courses taught by the same instructor. Controlling for class size and previous grades students rated teachers no differently, but course quality was significantly higher in

intensives. In a study of an intensive psychology course (termed immersion scheduling), a comparative study indicated 2-week duration intensive students rated evaluated course and instructors significantly higher (Richmond et al., 2015). Research on intensives suggest potential benefits in student performance and satisfaction if scalable.

Block mode

Block or 'One-Course-At-A-Time' mode consists of homogenously structured intensives studied in sequence, adopted as the primary format. Scholarly literature indicates block mode has been adopted mainly by relatively small-cohort (approximately 700-2500 students), private not-for-profit, secular liberal arts Colleges where most students live on campus. A form of block mode has been delivered at Colorado College from 1970 (Drake, 1973), Cornell College from 1978 (Vaughan & Carlson, 1992), Hiram College from 1934 to 1961 (Vaughan & Carlson, 1992), Tusculum College from 1994 (Artis & Overton, 2010) and more recently Quest University from inception (Bouw, 2013; Helfand, 2013) and in a tourism management degree at an unspecified UK-based university (Dixon & O'Gorman, 2019). Some older and larger Western Universities also offer block mode, such as Malmo University in Sweden, but the model is not homogenous and blocks vary greatly (Malmo University, 2019).

A ten-year evaluation at Colorado College provides a fascinating historical record of a major university change, and the empirical component shows promising results for block mode. Surveys undertaken in the inaugural and following year of implementation found 73% of staff were satisfied with the model and student satisfaction was above 90% (Drake, 1973; Heist & Taylor, 1979). Using a comparison university, results indicated no major declination of educational progress. Vaughan and Carlson (1992) summarise evaluative efforts at Cornell College. Staff and student surveys, along with interviews, undertaken immediately post-

introduction and up to ten years later, indicated increased perceived engagement and academic performance, and preparedness for graduate school. Both studies lacked baseline data, and rely on descriptive statistics and retrospective self-report surveys.

Artis and Overton (2010) analyse national attrition rates, finding Tusculum College had a comparatively high retention rate after controlling for institutional factors, but that the block model was not a significant predictor of retention when other Colleges offering block mode were included. Dixon and O'Gorman (2019) undertook an anonymous survey with nine lecturers who made the transition to block mode, who reported increased perceived focus and faster sense of accomplishment, but also increased time pressure and more rapid falling behind. Like Lutes and Davies (2018) study of intensives, some also reported potential threats to deep learning. Large sample studies with data pre and post introduction of block mode are absent.

With a VU to improvement: The VU block model. The block mode began at Victoria University, Melbourne, Australia (VU) partly in response to low quality and performance indicators. With just over 28, 000 students in 2019 VU is one of 6 dual-sector Universities in Australia, offering courses across all disciplines and Australian Qualification Framework levels (Australian Qualifications Framework Council, 2013). In a national context, VU caters for a high proportion of culturally and socio-economically diverse students (Department of Education, 2019).

The removal of Government caps on student places resulted in expansion of larger and more elite Universities, but relative stagnation and decline for VU leading to budget constraints (Victorian Auditor-General's Office, 2017, p. 38). VU scored low on some indicators of the national teaching and learning quality survey (QILT, 2017; see Table 10), ranking last for overall

experience, and low for learning resources, student support and teaching quality. VU does, however rank around middle of the nation for skills development and learner engagement.

Block mode at VU was introduced in a major overhaul of the first-year experience detailed in an internal 'white paper', the VU Transformation Agenda (2017), which states the goal is to achieve the highest first-year student satisfaction of any university in Victoria; in addition to furthering the university mission of providing excellent and accessible education. This initiative created a new non-disciplinary college with a focus on transitions pedagogy (Kift, 2015) and aligned professional development. Teachers were recruited to this college who had a track record of interest, quality and innovation in teaching. However, senior leaders central in the decision and implementation determined a more radical change was necessary:

> "*Internal discussions within VU called for a solution that moved beyond, or arguably, more wholeheartedly adopted, the frameworks and principles of first year transitions. What was required was an approach that, while incorporating these initiatives, would also position the university as legitimately agile, innovative, open and ready to deliver an educational experience suited to the 21s-century student"* (McCluskey et al., 2018a, pp. 6)

The senior leadership group undertook 'town hall' style meetings with staff to discuss the block concept in 2017. Senior leaders were greatly inspired by a TEDx West Vancouver talk from Prof. David Helfand, then of Quest University (Helfand, 2013), and visited both Quest and Colorado College to observe. Unlike Colorado College where a precondition of a new model was majority vote by Faculty in a secret ballot, the decision at VU was not subject to a vote.

Figure 1 visually contrasts the block and traditional mode. Time devoted to face-to-face and self-directed study are roughly equal, but reorganised. Advantages are similar to those claimed for intensives, blended and active learning, in that concentrating on a single unit at a time reduces scheduling and time management challenges, and increasing small-group active learning over lectures enables greater immersion, engagement, deeper connections with teacher and peers, and ultimately, learning. The block model was implemented in first-year units in 2018 (a small minority remained in traditional mode for course architecture/pathways).

The model consists of maximum 11 block per year, comprising two summer blocks and one winter block - 'off-season' blocks undertaken by fewer students - and eight core blocks replacing the previous core two-semesters. Each block is four weeks in duration with a blended/flipped curriculum design combining online learning and digital learning tools with three on-campus classes per week, each of three hours (referred to as a '3-by-3' schedule; 12 sessions in total). Lectures are removed, replaced with classes capped at 35 students per class, comprising small-group interactive learning in fixed class cohorts to encourage social engagement and support. Pre and post-class activities complement classes, which are mostly completed online but can also include experiential learning such as work placements and field trips. Complementary activities mainly comprise on-campus workshops that focus on core academic skills and future readiness (including employability).

[*Place Figure 1 around here]*

The model includes two short two-week breaks but otherwise blocks run continuously, and students can enroll in as many or few blocks as they wish. Assuming necessary blocks are offered in the desired periods, students can attain a qualification faster than previous (typically 24 blocks in a Bachelor degree). With this level of flexibility, study patterns are more complex

and take any conceivable form. However most students take eight blocks-per-year matching the two-semester model.

Design Principles and Processes. Consistent with literature on intensives and constructive alignment (Biggs, 1996), block design principles and processes stress active and small-group learning, backward design, immersion, and rapid feedback (Table 1). Key design features include a two-day turnaround on marking assessments, a low-stakes assessment in the first week of every block, a census period of 1 week (the date a student must withdraw or incur a Government loan), and removal of exams where accreditation requirements allow. Development of semi-automated assessment rubrics greatly supported accurate and efficient assessment feedback.

All block units undergo a five-step continuous design process (Figure 2). Teams comprising key academics, learning designers, librarians and students (often employed as staff), and a formal academic peer review of the unit design. It is noteworthy that while VU has a number of ongoing quality improvement process that intend to enhance teaching and learning, including reviewing unit design, the block project is the first time that *all* first-year units have undergone a central curriculum review process.

Implementation. Almost all staff members, departments and functions at the university were affected by the block project: implementation is large and complex. Resultantly there is no straightforward cost as it is difficult to quantify required staff time. However, a strategic (limited duration) expenditure was approved by the university Council to support implementation. This expanded two key central Departments: the Strategic Project Office (STO) that undertakes project management and Connected Learning (CL), a central teaching and learning department.

CL were central in the initiation and implementation of the block, and subsequently led the multi-departmental and disciplinary unit-design teams. Both before and during block mode, CL, in conjunction with Colleges, also led professional learning and curriculum development activities including continuing to offer a formal university teaching-focused qualification (Graduate Certificate in Tertiary Education). STO project-managed timelines and milestones of the roll-out. In sum, 156 first-year block units were designed and delivered in 2018. Many other aligned and ongoing functions of the university required substantial redevelopment.

Some results have been published, indicating an increase in pass rate and grade distribution and overall increase in retention (McCluskey et al., 2018b). An internal evaluation in 2018, primarily qualitative, comprising interviews and focus groups with leaders, staff and students, concluded the block was mostly successfully implemented (Ambler, 2018). The block has been deemed so successful the decision has been undertaken to transition the entire university to block mode, and continuous evolution of the block mode is central to the university strategic direction.

[*Place Figure 2 around here]*

[Place Table 1 around here]

Student Satisfaction

Student satisfaction was a key goal of the block model and has been the focus of both ongoing quality assurance activities and educational research since at least 1924 (Aleamoni, 1999). In fact “many 1000s of studies have been conducted on the validity and diagnostic usefulness of students’ evaluations of university teaching” (Marsh et al., 2019, p. 1), which also measure satisfaction with curriculum and are commonly abbreviated as SETs. Reviews conclude SETs have acceptable measurement characteristics, validity and usefulness (Marsh, 2007;

Spooren et al., 2013), and a number of biases including gender (Centra & Gaubatz, 2000), likeability (Feistauer & Richter, 2018), attractiveness (Felton et al., 2004), and temporal effects such as a mood (Zumbach & Funke, 2014) or providing cookies with the assessment (Hessler et al., 2018). Some consider relationships between satisfaction and student marks (or 'easiness', Felton et al., 2004) as evidence of bias, however we consider this relationship theoretically sound as higher student satisfaction should also enable higher participation and engagement, resulting in better marks (Richardson et al., 2012). Comprising Likert-type measures, SET-like surveys may also require complex statistical treatment (Kitto et al., 2019).

Many SETs also tap satisfaction with unit design, generally reflecting the concept of constructive alignment (Biggs, 1996). These include factors measuring 'clarity of objectives', 'appropriate workload', 'appropriate assessment' and 'clear goals and standards' (see Spooren, et al., 2013, Table 1, p. 606-607 for a short list). In Australia, the Federal Government regulator mandates the collection and use of student feedback but does not specify the form (Department of Education and Training, 2015, p. 11). Australian Universities and many others globally, employ varieties of SETs to measure student satisfaction. With many thousands of completion, SET like surveys reflect one of the most substantial forms of student voice.

Present Study

Studies of block mode are limited to small-scale and unique institutions, limiting the generalisability of findings to large-scale multidisciplinary Universities. VU may be the first university of this scale to make the transition. This study treats the introduction as a natural experiment (Morgan & Winship, 2015). Natural experimental logic and analyses are particularly suitable to educational settings and interventions, as experimenter control is often constrained due to ethical or practical reasons. In the place of randomising participants into conditions,

examining the changes that occur in long-standing institutional indicators, utilising comprehensive statistical controls, is a promising way to infer the effect of an educational intervention.

We compare the inaugural block cohort to counterparts from two prior years on satisfaction and academic performance. As Universities comprise structures (disciplines, units) and dynamic aspects (changing students, teachers), we apply comprehensive statistical controls in mixed effect cross-classified models (Raudenbush & Bryk, 2002). When presented with substantial analysis choices we report sensitivity tests. To ensure no students are being 'left behind' and to identify potential explanatory mechanisms, we undertake extensive moderation testing on the effect of the block. This includes testing all covariates, and a level two moderator of altered assessment design.

Hypotheses.

Original hypotheses were agreed (email correspondence, 03/12/2018) before analysis or data extraction, and pre-registered on the Open Science Framework (OSF), see (osf.io/knwv7). Small alterations to these hypotheses were subsequently undertaken, and final hypotheses are presented in Table 2. Hypotheses were altered for the following reasons. Firstly, we reorder hypotheses to be more consistent with the order of analyses: hypotheses on the validity and reliability of satisfaction scales are listed before hypotheses on the effect of the block. Secondly, and representing the largest changes, certain planned analyses were not undertaken due to time / resource constraints - all analysis of the qualitative (open text) comments were deemed beyond scope. We also add one further source of moderation at the unit-level in the form of altered assessment design, enabled by extracting data from a university curriculum monitoring system,

initially considered unsuitable for data extraction. Finally, we improve the grammar and specificity to avoid any confusion, such as replacing ‘over time’ with ‘block mode’.

[Place Table 2 around here]

Method

Measures

Data was drawn from institutional student enrolment, gradebook, satisfaction and the curriculum management systems at the university. Institutional data were extracted, merged and validated primarily by the second author. The project received ethical approval from the Victoria University Human Research Ethics Committee (HRE17-192). Satisfaction is measured on conclusion of every unit with the Satisfaction with Unit (SEU; often termed course satisfaction in the literature) and the Satisfaction with Teacher (SET) survey (Table 3). Each has six items with a 5-point Likert-type response range (item wordings are in the results section for ease of reference). We investigate validity of the SET and SEU in the preliminary results. Survey items changed in 2016 and for this reason, the present dataset begins then. Student marks represent the final, weighted sum of all summative assessment for a corresponding unit. Control variables comprise comprehensive indicators of student demographics, pathways, equity indicators, and discipline (see Appendix A, Table SM-A-1). We also derive a variable reflecting change in unit assessment in transition to block.

Procedure. Many measures used are collected as routine university operations. Administration of satisfaction surveys occurs via the online learning management system and student email address. Teachers likely vary in their promotion and framing of SET and SEU, which can influence response rates (see Thielsch et al., 2018), but we have no indicator of teacher recruitment approaches. Students can complete the SET for as many teachers as they had

during that unit and period. Where students rated more than one teacher, we aggregate across teachers at the student-level and take the teacher ID based on the first teacher listed in the system (19.65% of SET respondents, see *Sample*).

Analysis Strategy

As a natural experiment, the control group comprises two cohorts of first-year pre-block Bachelor students (years 2016 and 2017), and intervention group the initial cohort of first-year Bachelor block students (2018); with block mode dummy-coded. We utilise statistical controls to decompose the effects of the block from changing student characteristics, random intercepts to capture the effect of different units, teachers and students, and a linear and quadratic time variable to account for pre-existing trends or seasonality (Supplementary Materials, Appendix A). As stated by Marsh, "Global or "overall" ratings cannot adequately represent the multidimensionality of teaching" (Marsh, 2007, p. 327). Based on the unique content of each item, and potentially variable effect of the block on aspects of satisfaction, we analyse items separately.

Statistical analysis. The dataset includes hierarchies and crossed cases. Students can rate more than one unit, some teachers are present in both or only one model (block or traditional), and students are nested in units and teachers. As such we utilise cross-classified linear mixed effects models with random intercepts for unit, teacher and student (Bates et al., 2007). The basic underlying equation is:

$$Y = \alpha + \beta_{time} + \beta_{time^2} + \beta_{block} + \mathrm{B}X,$$

where BX is a matrix of pre-treatment covariates. Random components of the cross-classification model formula are lengthy so this equation presents only the fixed components. Failing to account for nested structures of the data can lead to erroneous conclusions (Raudenbush & Byrk,

2002), and incorporating them in models strengthens generalisability. Where presented with substantive analysis or sampling choices we undertake and report sensitivity tests. Statistical significance criteria is that the 95% confidence interval does not include zero, except in the case of ANOVAs comparing marginal effects, where we adopt $p<.05$ of the F statistic as the criteria.

***Missing Data*.** Missing data patterns were examined (see Table SM-D-3 in Supplementary Materials). Almost all students who began a satisfaction survey completed all twelve items (99.97%), however average response rate per unit was fairly low at 19.4% (*SD*=7.35) [20.04, 18.78]. Teacher ID values were unavailable without completion of the SET as the university has no central dataset pairing a given student and delivery period with certain teacher(s). Hence models that include a random intercept for teacher exclude non-respondents.

Similar response rates have been reported in other student satisfactions studies (Berk, 2012). Online implementation of SETs as in the present case has been associated with lower response rates than paper-based, and non-response is driven by a range of personal reasons (see Berk, 2012, p. 100; Nulty, 2008). Thielsch et al., (2008) found student demographics and opportunity costs do not predict SET participation, while perceived salience, frequency of peer participation, and a non-domain specific willingness to undertake surveys do. Control variables were sourced via enrolment processes and had no or low (<3%) missing values, with exception of Australian Tertiary Admission Rank[1] scores (22.25% missing). An increasing number of enrolments are direct entrants who may not have or report an ATAR; an intended consequence of expanding university access (Bradley et al., 2008).

***Treatment of Missing Data*.** Green (2016) reviewed missing data handling in higher education studies. Of 143 articles only 40% even reported the presence of missing values, and in studies that treated missing data 8 (20%) used listwise deletion and 10 (25%) expectation

maximisation algorithms. We elected to use the bootstrapped expectation maximisation algorithm in Amelia (Honaker et al., 2010), suitable for cross-sectional time series data (see Zhang, 2016 for applied examples), and considered superior to mean or random normal replacement (Moritz et al., 2015). Imputed valued were restricted to the original variable range, and as each imputation varies we produce and fit models across 100 datasets, pooling the standard errors in results. We do not impute missing teacher ID variables. Imputed and original distributions are in Supplementary Materials, Figure SM-D-1.

Preliminary Analyses (hypotheses 1a and 1b). First we examine validity and psychometric qualities of the SET and SEU. Content validity is assessed via comparison with established SET scales and the Australian National Student Experience Survey or SES (QILT, 2017), convergent validity via relationships with marks and relevant SES constructs, and psychometric characteristics via confirmatory factor analysis with discriminant validity through nested model comparison across a one and two-factor solution (Bagozzi et al., 1991).

Primary Analyses (hypothesis 2). We first test intra-class correlations (ICCs) that capture variance explained by grouping factors. Accounting for covariances between outcome variables has been shown to impact key results only rarely (Jackman, 2009), hence we fit separate models for each outcome variable. Mixed effect cross-classified models were fit to examine the effect of the block on satisfaction indicators and performance.

Sensitivity Analyses (hypothesis 2a) Forking paths in sampling and analysis led to sensitivity tests. Sampling choices were: 1. Including or excluding non-participating students (i.e. students with almost no participation the unit's assessments); and 2. Restricting the sample to students who completed a SET (with a corresponding teacher ID enabling a random intercept). Kitto and colleagues (2018) discuss the complexity of analysing Likert data in SET surveys.

Partly as linear models assume a continuous underlying scale they argue linear models are inappropriate. Resultantly, we also fit ordinal regression cumulative link mixed models (see Christensen, 2019) as a sensitivity test. These models do not assume equal distances between scale anchor points and also include random intercepts for unit, student and teacher.

Moderation analyses (hypothesis 2b). As an exploratory study of a new model we undertake extensive moderation testing. Given limited knowledge of block mode, including whether certain student subgroups may be more or less satisfied or successful, we test every control variable as a moderator of the treatment effect. We fit separate models testing each potential moderator and evaluate significance of the interaction term. For multi-categorical variables, an ANOVA tests for differences across marginal effects.

Software. A large number of R packages were utilised. ICCs were calculated with *Multilevel* (version 2.6), CFAs were fit in *Lavaan* (version 0.6-5), *MICE* examined missing value patterns (version 3.6.0), and *Amelia* (version 1.7.5) generated imputations. Mixed effects linear models were estimated using lmer in *lme4* (version 3.1-141), with MIcombine in *mitools* pooling results (version 2.4). Interaction effects for continuous moderators are plotted with *interplot* (version 0.2.1; see Solt & Hu, 2018). An ANOVA using the D1 method was calculated for multi-categorical marginal effects using mi.anova within *miceadds* (version 3.7.6, see Robitzsch, 2019). Estimated marginal effect confidence intervals were produced using deltaMethod in *car* (version 3.0-5). Ordinal regressions were fitted using clmm in *ordinal* (version 4-25). Example model code are in Table SM-A-3 in Supplementary Materials.

Results

Sample

To ensure reproducibility of the sample and analysis, we provide the filtering process and inclusion criteria used to reach the final sample in Table A-SM-3 in Supplementary Materials. In short, inclusion criteria required a unit to have satisfaction data pre and post block mode. 157 units had enrolments in the eight core block delivery periods in 2018. When examined, 18 of these units had no enrolments in 2016 or 2017, as they are likely newly created units. Two additional units had no completions of the SET survey in either pre block or post block periods.

Results are therefore based on a sample of 136 units with satisfaction information pre and post block mode, comprising a total of 15,989 observations (n = 5775 block, n = 10,214 traditional), including n distinct students = 6325 (n=2504 block, n=3498 traditional, n=323 both), and n distinct teachers = 616 (n=131 block, n=373 traditional, and n=112 both). Of the distinct students, 3624 or 57.29% rated more than one unit in the sample (M = 2.51, SD = 1.86), and 19.65% rated more than 1 teacher in a given period. The dataset for marks includes n observations = 82,031, from n = 17,676 distinct students. A small number of students also rated the same unit twice (n repeat unit ratings = 132), indicating they likely did not pass and repeated the unit. Some students also undertook block and traditional units in overlapping delivery periods in 2018 (n multimodal students=719 / 11.3%); we include a flag for these as a covariate in all models.

Response Rates. Unit response rate for any SET or SEU item was M=19.41 [20.04, 18.78] (SD=7.35). Response rates increased marginally in the block mode (M= 20.88 [18.93, 22.83], SD= 11.61) compared with traditional mode (M=18.58 [17.23, 19.93], SD=8.05), and also varied by discipline and year, with a notable increase in block mode for Engineering and Science (Supplementary materials, Table SM-D-2).

[Place Table 3 around here]

Descriptive statistics

Original and imputed variable distributions are in Supplementary Materials (Figure SM-D-1). Overall, students are satisfied with scores negatively skewed. This high satisfaction seems to contrast with the ranking of the university in the national scale. Simple inferential mean comparisons indicate marks increased substantially in block mode but most SEU items did not, except for increased SEU5 and decreased SEU6 (Supplementary Materials Table SM-D-3). SET on the other hand showed increases in every indicator.

Assessment design changes are in Figure 3. The largest changes in block mode were decreased use of exams and essays and increased use of tests, presentations, and class exercises. The 136 units show a roughly normal distribution on the total assessment change variable, with exception of a concentration of units that changed very little or not at all.

[Place Figure 3 around here]

Preliminary analysis (hypotheses 1a and 1b)

Hypotheses 1a and 1b were confirmed: the SET and SEU scales demonstrate sufficient content, convergent and psychometric validity to proceed with primary analyses (see Supplementary Materials, Table SM-A-1). Item content aligned with more established SET-like scales and the national student satisfaction measure (SES), convergent validity was present in relationships with marks, and also significant but mostly small correlations with some relevant SES constructs. Confirmatory factor analyses indicated acceptable and superior fit for a two-factor solution. However, the negative skew of the distribution suggests range-restriction.

Intra-class correlations. To test nested structures we calculate intra-class correlations or ICCs (Supplementary Materials, Table SM-E-1) and compare random-intercept to intercept-only models (Bliese, 2006, p. 55; Hox et al., 2017, p. 13). Across satisfaction items, unit explained 6-

8% of variance, student 16-23% of variance, and teacher 11-23% of variance. For marks, 8% of variance was explained by unit, 65% by person, and 19% by teacher. In all cases random intercept models had significantly better fit.

Primary Analyses (Hypotheses 2, 2a, 2b)

Treatment main effects (the effect of the block) with sensitivity tests are presented in Table 4 for marks and Table 6 for satisfaction. Significant categorical moderators are in Tables 5, 7 and 8, with continuous moderators presented in a panel of plots in Figure 4. As control variables are not of central interest, fixed effects for all model parameters are reported in Supplementary Materials, Appendix A. β denotes effects in standard deviation units and B in unstandardised.

Hypothesis 2 was partially confirmed. The block had a very large positive effect on marks. Including a random intercept for teacher made little difference to the effect on mark (Model M-S-2), as did excluding non-participating students (Model M-S-1), so we adopt the full sample model as the primary result: B(se) [95CI] = 11.11(0.34) [10.44, 11.79]. The block also increased five of six teaching satisfaction indicators, but decreased four of six unit satisfaction indicators. Satisfaction effects are small and often border zero, the largest was a decrease in perceived reasonable workload (SEU6: β(se) [95CI] = -0.24(0.05) [-0.34, -0.15]).

Hypothesis 2a was partially confirmed for satisfaction and marks. Ordinal regressions produced fairly consistent results but with greater declines in unit satisfaction and a number of teaching satisfaction indicators now non-significant. The effect on marks reduced marginally when non-participating students were excluded (B(se) [95CI] = 9.64(0.28) [9.09, 10.19]).

Hypothesis 2b was also partially confirmed. As moderations tests are exploratory and numerous only key trends are discussed. Broadly, treatment effects were robust and rarely did

marginal effects reverse in direction. Like the main effects, satisfaction moderators were small. Certain disciplines, older students and VTAC applicants showed greater declines in unit satisfaction and smaller gains in teaching satisfaction. Somewhat counter-intuitively, more academically accomplished (higher ATAR) students showed a greater decline in perceived reasonable workload. The greatest gains in marks were for younger, lower ATAR, non-English speaking and low socio-economic students.

Changed assessment design offset a small but significant amount of decline in unit satisfaction, marginally increased marks, but also led to slightly lower perceived ratings on two teaching indicators. As the distinction between tests and exams were unclear, a further test was undertaken, in which any decrease in examinations with a corresponding increase in tests were coded as no change. This decreased the mean change in assessment design, but did not change the moderation term.

[Place Table 4 around here]

[Place Table 5 around here]

[Place Table 6 around here]

[Place Table 7 around here]

[Place Table 8 around here]

[Place Figure 4 around here]

Discussion

This study is the first empirical investigation of introducing block mode on student satisfaction and performance in a large-scale university with baseline data. With expanding access, it is imperative Universities provide the best possible opportunities for *all* students, including the 'new majority' (Keller, 2001). A way to achieve this may be to reform the mode of study. Research on active and blended learning show promise (Dawson & Dawson, 2018; Vo et al., 2017), as do intensives (Wlodkowski, 2003; Austin & Gustasfon, 2006). The block combines aspects of these in a primary format, but has so far been applied and evaluated in small, unique institutions, with studies of introduction lacking baseline data (Heist & Taylor, 1979; Vaughan & Carlson, 1992). Introducing a block model to a university is a large undertaking, and to ensure effectiveness and in keeping with the ethos of inquiry, any such innovation must also be evidence-based.

The most striking effect of introducing a block mode was lifting assessment results. Students are predicted to achieve over ten marks higher in block than traditional mode – an entire grade category (for example, a 'credit' is predicted to move to a 'distinction'). The effect on marks was robust, remaining positive for all moderators tested, and changing unit assessment design only increased this effect further. Importantly, the effect was largest for younger, non-English speaking background, less prior education, lower academic performance and socio-economic status students: key equity groups in an expanding university system (Baik et al., 2019). However, the effect varied by discipline with greatest gains in business and smallest in arts and education. Factors that reduced the effect on marks were studying part time and age. At roughly half the overall effect size, part time and older students will require further attention.

Improved academic outcomes align the block with some studies of intensives, active and blended learning (Austin & Gustafson, 2006; Dawson & Dawson, 2018; Vo et al., 2017), which also found increases in assessment results. The standardised effect of the block on assessment (Model M-S2; 0.45 (0.01), [0.43, 0.48]) is somewhat comparable to the ranking of meta-effects on student achievement reported in Schneider & Preckel (2017). In this ranking, the block falls between ranks 29 (extracurricular training) and 37 (organisation as a learning strategy). We theorise two main drivers of this effect. One is the removal of competing schedules and some planning, task and time management challenges that come with studying units in parallel. The other is increased engagement in the new model. However, we were not able to test these two mechanisms directly as consistent measures of academic time management ability and engagement were not available pre and post block.

Satisfaction results paint a more complex and mixed picture. Descriptive statistics indicate most first-year students were satisfied both before and in block mode, consistent with many SET studies (Avery et al., 2006; Fike et al., 2010; He & Freeman, 2020). While the satisfaction surveys demonstrated content, psychometric and some convergent validity, possible range restriction may have dampened effects. It is recommended that a larger range be tested and pending psychometric qualities, be considered for adoption moving forward. Similarly, as pre-block levels of satisfaction are already close to the top of the range, potential gains of introducing the block model are more limited, with far more potential downward movement.

Overall, effects of the block on satisfaction were small. The largest is less than a third of a standard deviation (SEU6, $\beta = -0.24$) and many effects approach zero in 95% confidence intervals. For context, Kuscsera and Zimmaro found 9-week intensives had a positive effect on course satisfaction over traditional length units ($\beta = 0.22$, $p<.001$, p. 66), and temporal effects

have been similar (Hessler et al., 2018, p. β=0.21, *p*<.03 p. 1069). The block increased five of six teaching satisfaction indicators but decreased four of six unit design satisfaction indicators.

In terms of teaching, block students were more satisfied overall, with helpful feedback, interestingness, efforts to understand difficult and supporting academic motivation. Ratings of being good at explaining content did not change, consistent with the removal of lectures and facilitator-role in block teaching (Helfand, 2013). In terms of unit design, declines were evident in ratings of overall unit quality, learning activities, learning resources being relevant and up-to-date, and the largest - a reasonable workload. The largest effect seems consistent with the risk of exhaustion in intensives (Male et al., 2017). Ordinal model findings were broadly consistent but enlarged unit satisfaction declines and suggested only one improved teaching satisfaction indicator (efforts to understand difficulties).

Declines in unit quality, learning activities and resources are surprising given the focus in block mode on small-group learning, and the investment in unit redesign, which involved improving learning resources, especially online and interactive content. These findings may reflect teething issues in such a rapid and large-scale change project, and future studies are required to determine whether satisfaction improves with continued bedding down of the model. Given the initially high levels of satisfaction, indicators had much farther to fall than rise, and the lack of change may reflect the investment in central unit design supporting implementation.

Moderation testing revealed effects on satisfaction are fairly robust across student sub-groups: most interaction terms were non-significant and rarely did marginal effects reverse in sign. Discipline, pathway indicators and unit assessment design change were fairly consistent, significant but small moderators of satisfaction. As moderation testing was exploratory and numerous, every finding is not discussed in depth. Notably larger declines in unit satisfaction

were evident in the disciplines of health and law. As the university formed a discipline-free College for first-year (with all disciplines working closely together), this finding likely reflects content-related factors and possibly varied implementation, as design teams worked by discipline. Lutes and Davis (2018), in a survey of a small sample of academics teaching intensives, found most retained the same content in intensive mode, but recursive-based units heavy in reading and writing were more not be as well suited to the mode. In short, we thoroughly searched for, but could not find, student subgroups that were greatly more or less satisfied in block than their counterparts from prior years.

Interestingly, and consistent with guidelines on intensives, units that altered assessment design offset some decline in satisfaction. Beyond the likely direct effect of altering assessment to make it more viable in the block, it is also possible this variable reflects an implementation check, and measures the engagement of the unit and associated staff with the block redesign process. Paradoxically, higher ATAR students found the workload somewhat more unreasonable in block than lower ATAR students, although the effect is small. This effect is difficult to explain. Notably, prior education levels did not moderate satisfaction, suggesting that students with university experience did not demonstrate a preference for the traditional mode. By and large both the effects of block on satisfaction, and their moderators, were small.

The present study has some strengths over prior literature on intensives and block, but also several limitations. Strengths include the availability of pre-intervention/baseline measures that form a natural non-matched time series, with the use of comprehensive controls, a large sample, extensive sensitivity and moderation testing, and sophisticated statistical models suitable to the structure of the data. Limitations include only a year of block delivery, low average response rate per unit for satisfaction measures, and lack of other readily available and reliable

data sources to test implementation fidelity or explanatory mechanisms. Student time management ability and engagement (Walsh & Risquez, 2020), richer indicators of assessment design such as use and qualities of rubrics, student product, and content indicators such as volume of reading and use of long slide decks in classes versus uptake of small-group active learning, would have been ideal.

There are also many more sources of meaningful information not included the present analysis that future studies may consider. The block model affected almost every aspect of the university. Perhaps the most notable absence is information from Faculty, but also broader indicators such as student open-ended comments in surveys, other indicators of learning outside of the final assessment result of a unit, and longitudinal outcomes, which all require analysis. Perhaps the most significant limitation is that it is simply too early to make conclusions about the block mode, with only a single year of implementation. While investigating the inaugural cohort is necessary, when considering an evaluative plan for introducing block mode at Colorado College administrators decided that “the [block] plan would be at least a two-year experiment after which the faculty would decide to abandon, modify or go on” (Heist & Taylor, 1979, p. 32). Further, ongoing studies with subsequent block cohorts are required.

With the expansion of university study in Australia and globally, Universities must strive to preserve the most valuable elements of higher education while innovating to better support a growing number and diversity of students. The block model reflects one of the most radical reformations and must be accompanied by careful empirical investigation. The block had a dramatic, robust positive effect on marks, that benefitted sub-groups generally at risk of poorer performance and important in an expanding sector. However, the block also had small and mixed effects on satisfaction. With ongoing development of the model, it is hoped improvements in

student performance and teaching satisfaction are cemented, and unit satisfaction indicators will match and surpass prior levels.

Endnotes. 1. Higher education terminology varies. In this paper ‘unit’ refers to a single subject, ‘course’ refers to all study comprising an entire qualification, ‘tutorial’ refers to classroom-based learning, which typically takes the form of a seminar or workshop. For further detail see the Australian Qualifications Framework. 2. Australian Tertiary Admission Rank (ATAR), is the culminating academic ranking score gained in secondary school, used to prioritise university placements in Australia for those applying through that system. Direct applicants to university who have no ATAR score have risen in line with expansion of the sector. 3. Dual sector is a term in Australian Higher Education indicating the institution offers courses in both Vocational Education and Training, and at university-level. Further information on Australian Higher Education is in Supplementary Materials, section Appendix C.

References

Aditomo, A., Goodyear, P., Bliuc, A. M., & Ellis, R. A. (2013). Inquiry-based learning in higher education: principal forms, educational objectives, and disciplinary variations. *Studies in Higher Education*, *38*(9), 1239-1258. http://doi.org/10.1080/03075079.2011.616584

Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education*, *13*(2), 153-166.

Ambler, T. (2018). *Evaluation of the first year model: Insights from year one*. [Internal Document (not publicly available)]. Victoria University, Melbourne, Australia

Artis, R. C. & Overton, M. B. (2010). Tusculum College's block calendar system: Analysis of the impact of the block calendar system on retention of first-year Tusculum College students. https://ir.vanderbilt.edu/bitstream/handle/1803/5199/Capstone-%202010-%20Artis,%20Overton.pdf?sequence=1

Austin, A. M., & Gustafson, L. (2006). Impact of course length on student learning. *Journal of Economics and Finance Education*, *5*(1), 26-37.

Australian Qualifications Framework Council (2013). *Australian qualifications framework: Second edition*. https://www.aqf.edu.au/sites/aqf/files/aqf-2nd-edition-january-2013.pdf

Avery, R. J., Bryan, W. K., Mathios, A., Kang, H., & Bell, D. (2006). Electronic course evaluations: Does an online delivery system influence student evaluations? *Journal of Economic Education*, *37*, 21–37.

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative Science Quarterly*, 421-458.

Baik, C., Naylor, R., & Arkoudis, S. (2015). *The first year experience in Australian Universities: Findings from two decades, 1994-2014*. Melbourne Centre for the Study of Higher Education. https://melbourne-cshe.unimelb.edu.au/__data/assets/pdf_file/0016/1513123/FYE-2014-FULL-report-FINAL-web.pdf

Baik, C., Naylor, R., Arkoudis, S., & Dabrowski, A. (2019). Examining the experiences of first-year students with low tertiary admission scores in Australian universities. *Studies in Higher Education*, *44*(3), 526-538. http:/doi.org/10.1080/03075079.2017.1383376

Barnett R. (2019). Re-valuing the university: An ecological approach. In: P. Gibbs, J. Jameson & A. Elwick (eds) *Values of the university in a time of uncertainty* (pp. 47-59). Springer

Bates, D., Sarkar, D., Bates, M. D., & Matrix, L. (2007). The lme4 package. *R package, 2*(1), 74. ftp://ftp.uni-bayreuth.de/pub/math/statlib/R/CRAN/doc/packages/lme4.pdf

Berk, R. A. (2012). Top 20 strategies to increase the online response rates of student rating scales. *International Journal of Technology in Teaching & Learning*, *8*(2), 98-107.

Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education*, *26*(1), 87-122. http://doi.org/10.1007/s12528-013-9077-3

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, *32*(3), 347-364.

Bliese, P. (2006). *Multilevel modeling in R (2.2): A brief introduction to R, the multilevel package and the nlme package*. http://ftp.uni bayreuth.de/math/statlib/R/CRAN/doc/contrib/Bliese_Multilevel.pdf

Bouw, B. (2013, October 7). A Capital Quest. *BCBusiness*.

https://www.bcbusiness.ca/a-capital-quest

Bradley, D., Noonan, P., Nugent, H., & Scales, B. (2008). *Review of Australian higher education: Final report.* Department of Education and Training. http://hdl.voced.edu.au/10707/44384

Burton, S., & Nesbit, P. L. (2008). Block or traditional? An analysis of student choice of teaching format. *Journal of Management & Organization, 14*, 4-19.

Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching?. *The Journal of Higher Education*, *71*(1), 17-33.

Christensen, C. M., & Eyring, H. J. (2011). *The innovative University: Changing the DNA of higher education from the inside out*. John Wiley & Sons

Christensen, R. H. B. (2019). *A tutorial on fitting cumulative link mixed models with clmm2 from the ordinal package*. ftp://ftp.ussg.indiana.edu/pub/CRAN/web/packages/ordinal/vignettes/clmm2_tutorial.pdf

Cobian, W. J., & Newberry, C. R. (2009). The effects of systematic changes to a business course over a three year period. *Delta Pi Epsilon Journal*, *51*(2), 76–85.

Colorado College (2019). *Welcome to Colorado College*. https://www.coloradocollege.edu/basics/welcome/

Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, *69*(9), 970-977.

Daniel, E. (2000). A review of time-shortened courses across disciplines. *College Student Journal*, *34*(2), 298-309.

Davies, W. M. (2006). Intensive teaching formats: A review. *Issues in Educational Research*, *16*(1), 1-20.

Dawkins, J. S. (1987). *Higher education: A policy discussion paper*. Australian Government Publishing Service. https://www.voced.edu.au/content/ngv%3A1033

Dawson, P., & Dawson, S. L. (2018). Sharing successes and hiding failures: 'Reporting bias' in learning and teaching research. *Studies in Higher Education*, *43*(8), 1405-1416.

Department of Education and Training (2019). *Selected higher education statistics, 2018, section 11 equity groups*. https://docs.education.gov.au/node/53025

Department of Education and Training (2015). *Higher education standards framework*. https://www.education.gov.au/standards-0

Dixon, L., & O'Gorman, V. (2019). 'Block teaching' – exploring lecturers' perceptions of intensive models of delivery in the context of undergraduate education. *Journal of Further and Higher Education.* http://doi.org/10.1080/0309877X.2018.1564024

Drake, G. A. (1973). An analysis of the Colorado College plan. *Critique*, *5*(2), 1-4.

Enders, C. K., and D. L. Bandalos. 2001. The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 8*(3), 430–57.

Fike, D. S., Doyle, D. J., & Connelly, R. J. (2010). Online vs. paper evaluations of faculty: When less is just as good. *Journal of Effective Teaching*, *10*(2), 42-54.

Feistauer, D., & Richter, T. (2018). Validity of students' evaluations of teaching: Biasing effects of likability and prior subject interest. *Studies in Educational Evaluation, 59*, 168-178.

Felton J, Mitchell J, Stinson M. (2004). Web-based student evaluations of professors: the relations between perceived quality, easiness and sexiness. *Assessment and Evaluation in Higher Education, 29*(1), 91–108. http://doi.org/ 10.1080/0260293032000158180

Froyd, J., & Simpson, N. (2008). Student-centered learning addressing faculty questions about student centered learning. *Course, Curriculum, Labor, and Improvement Conference, 30*(11), 1-11). http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.738.1395&rep=rep1&type=pdf

Fulton, K. (2012). Upside down and inside out: Flip your classroom to improve student learning. *Learning & Leading with Technology*, 39(8), 12-17.

Gauci, S. A., Dantas, A. M., Williams, D. A., & Kemm, R. E. (2009). Promoting student-centered active learning in lectures with a personal response system. *Advances in Physiology Education*, 33(1), 60-71. http://doi.org/ 10.1152/advan.00109.2007

Green, T. (2016). A methodological review of structural equation modelling in higher education research. *Studies in Higher Education, 41*(12), 2125-2155. http://doi.org/10.1080/03075079.2015.1021670

Gleason, N. W. (2018). *Higher education in the era of the fourth industrial revolution*. Palgrave Macmillan

Gold, A., Olin, N., & Wang, A. (2018, January 1). *What is the Delta method?* https://cran.r-project.org/web/packages/modmarg/vignettes/delta-method.html

Gumport, P. J., Iannozzi, M., Shaman, S., & Zemsky, R. (1997). *Trends in United States higher education from massification to post massification*. National Center for Postsecondary Improvement. Stanford University, School of Education

Harackiewicz, J. M., & Priniski, S. J. (2018). Improving student outcomes in higher education: The science of targeted intervention. *Annual Review of Psychology, 69*, 409-436.

Harvey, M., Power, M., & Wilson, M. (2017). A review of intensive mode of delivery and

science subjects in Australian universities. *Journal of Biological Eduation, 51*(3), 315-325.

Heist, P., & Taylor, M. F. (1979). *The block plan: A preliminary report on a ten-year evaluation of the Colorado College block plan format for intensive study.* The Colorado College

Helfand, D., (2013). *Designing a university for the new millennium* [Video]. TEDx Conferences (West Vancouver). https://www.youtube.com/watch?v=DZQe73IXZtU

Helfand, D. (2016). Mucking about in the mess: Research-based education at Quest University Canada. *Council on Undergraduate Research Quarterly, 36*(3), 28-33.

He, J., & Freeman, L. (2020). Can we trust teaching evaluations when response rates are not high? Implications from a Monte Carlo simulation. *Studies in Higher Education.* https://doi.org/10.1080/03075079.2019.1711046

Hessler, M., Pöpping, D. M., Hollstein, H., Ohlenburg, H., Arnemann, P. H., Massoth, C., ... & Wenk, M. (2018). Availability of cookies during an academic course session affects evaluation of teaching. *Medical Education*, *52*(10), 1064-1072.

Honaker, J., King, G., Blackwell, M., & Blackwell, M. M. (2010). *Package 'Amelia'*. http://kambing.ui.ac.id/cran/web/packages/Amelia/Amelia.pdf

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel Analysis: Techniques and Applications*. Routledge

Hsieh, P., Acee, T., Chung, W. H., Hsieh, Y. P., Kim, H., Thomas, G. D., ... & Robinson, D. H. (2005). Is educational intervention research on the decline? *Journal of Educational Psychology*, *97*(4), 523-534.

Huxham, M. (2005). Learning in lectures: Do 'interactive windows' help? *Active Learning in*

Higher Education, 6(1), 17-31. http://doi.org/%2010.1177/1469787405049943

Jansen, E. P. (2004). The influence of the curriculum organization on study progress in higher education. *Higher Education, 47*(4), 411-435. http://doi.org/10.1023/B:HIGH.0000020868.39084.21

Jackman, S. (2009). *Bayesian analysis for the social sciences.* Wiley

Keller, G. (2001). The new demographics of higher education. *The Review of Higher Education, 24*(3), 219-235. http://doi.org/ 10.1353/rhe.2001.0004

Kift, S. (2015). A decade of transition pedagogy: A quantum leap in conceptualising the first-year experience. *HERDSA Review of Higher Education, 2*, 51-86.

King, A. M., Mayer, C., Barrie, M., Greenberger, S., & Way, D. P. (2018). Replacing lectures with small groups: the impact of flipping the residency conference day. *Western Journal of Emergency Medicine, 19*(1), 11. http://doi.org/ 10.5811/westjem.2017.10.35235

Kitto, K., Williams, C., & Alderman, L. (2019). Beyond average: Contemporary statistical techniques for analysing student evaluations of teaching. *Assessment & Evaluation in Higher Education, 44*(3), 338-360. http://doi.org/10.1080/02602938.2018.1506909

Kucsera, J. V., & Zimmaro, D. M. (2010). Comparing the effectiveness of intensive and traditional courses. *College Teaching, 58*(2), 62-68. http://doi.org/10.1080/87567550903583769

Lutes, L., & Davies, R. (2013). Comparing the rigor of compressed format courses to their regular semester counterparts. *Innovative Higher Education*, 38(1), 19-29.

Lutes, L., & Davies, R. (2018). Comparison of workload for University core courses taught in regular semester and time-compressed term formats. *Education Sciences, 8*(1), 34-46.

Male, S., Baillie, C., Hancock, P., Leggoe, J., MacNish, C., & Crispin, S., Alam, F. (2016). *Intensive mode teaching guide.* Office for Learning and Teaching. http://www.ecm.uwa.edu.au/__data/assets/pdf_file /0005/2886008/Intensive-Mode-Teaching-GuideDec_21_2016.pdf

Malmo University, About Us (2019). https://mau.se/en/about-us/

Manicas, P. (2007). Globalization and higher education. In G. Ritzer (Ed.)., *The Blackwell Companion to Globalization*, (pp. 461–477). Blackwell Publishing

Marginson, S., & Considine, M. (2000). *The enterprise university :Ppower, governance and reinvention in Australia*. Cambridge University Press

Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In P. P. Raymond (Ed.)., *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective* (pp. 319-383). Springer

Marsh, H. W., Dickie, T., & Pfeiffer, M. (2019). A tale of two quests: The (almost) non-overlapping research literatures on students' evaluations of secondary-school and university teachers. *Contemporary Educational Psychology*, *58*, 1-18. http://doi.org/10.1016/j.cedpsych.2019.01.011

Marques, J. (2012). The dynamics of accelerated learning. *Business Education & Accreditation*, *4*(1), 101-112.

McCluskey, T., Weldon, J. & Smallridge, A. (2018). New kids on the block: Results of a first year college initiative. *Student Success.* https://unistars.org/papers/STARS2019/05A.pdf

McCluskey, T., Weldon, J., & Smallridge, A. (2019). Rebuilding the first year experience,

one block at a time. *Student Success*, *10*(1), 1-16. http://doi.org/ 10.5204/ssj.v10i1.1048

Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record*, *115*(3), 1-47.

Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference*. Cambridge University Press

Moritz, S., Sardá, A., Bartz-Beielstein, T., Zaefferer, M., & Stork, J. (2015). Comparison of different methods for univariate time series imputation in R. *arXiv preprint.* http://arxiv.org/abs/1510.03924

Naylor, R., Baik, C. & Arkoudis, S. (2018). Identifying attrition risk based on the first year experience. *Higher Education Research & Development*, *37*(2), 328-342. http://doi.org/ 10.1080/07294360.2017.1370438

Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: what can be done? *Assessment & Evaluation in Higher Education*, *33*(3), 301-314.

Organisation for Economic Co-operation and Development [OECD] (1987). *Universities under scrutiny*. ISBN: 9264129227

Organisation for Economic Co-operation and Development [OECD] (2008). *Higher education to 2030: Demography*. https://www.oecd.org/education/skills-beyond-school/highereducationto2030vol1demography.htm

Organisation for Economic Co-operation and Development [OECD] (2019). *Education at a glance 2019 – data and methodology (indicator C3)*. https://www.oecd.org/education/education-at-a-glance/educationataglance2019-dataandmethodology.htm

Quality Indicators of Learning and Teaching [QILT] (2017). *Student Experience Survey, 2017*

National Report. https://www.qilt.edu.au/docs/default-source/ses/ses2018/2018-ses-national-report75e58791b1e86477b58fff00006709da.pdf?sfvrsn=d733e33c_4

Quest University (2019). *The Block Plan*. Available at: https://questu.ca/academics/the-block-plan/

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd Ed). Sage

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353-386.

Robitzsch, A. (2019). *mi.anova.* https://www.rdocumentation.org/packages/miceadds/versions/3.5-14/topics/mi.anova

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

Scott, P. (2003). Attributes of high-quality accelerated courses. *New Directions for Adult and Continuing Education*, *97*, 29–38. https://doi.org/10.1002/ace.86

Scott, P. A. & Conrad, C. F. (1992). A critique of intensive courses and an agenda for research. *Higher Education: Handbook of Theory and Research*, *8*, 411-445.

Social Research Centre (2018). *Student Experience Survey 2018 Data Dictionary*

Solt, F., & Hu, Y. (2018). Interplot: Plot the Effects of Variables in Interaction Terms. https://cran.r-project.org/web/packages/interplot/vignettes/interplot-vignette.html

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, *83*(4), 598-642.

Tight, M. (2019). Systematic reviews and meta-analyses of higher education research. *European Journal of Higher Education, 9*(2), 133-152. http://doi.org/10.1080/21568235.2018.1541752

Thielsch, M. T., Brinkmöller, B., & Forthmann, B. (2018). Reasons for responding in student evaluation of teaching. *Studies in Educational Evaluation, 56*, 189-196.

Trow, M. A., & Burrage, M. (2010). *Twentieth-Century Higher Education : Elite to Mass to Universal*. Johns Hopkins University Press

van Ooijen-van der Linden L, van der Smagt MJ, te Pas SF and Woertman L (2018) A signal detection approach in a multiple cohort study: Different admission tools uniquely select different successful students. *Frontiers in Education, 3(*30), 1-11. http://doi.org/10.3389/feduc.2018.00030

Vaughan, C., & Carlson, C. (1992). Teaching and learning one-course-at-a-time. *Innovative Higher Education, 16*(4), 263-276. http://doi.org/10.1007/BF00917812

Victorian Auditor-General's Office (2017). *Results of 2017 Audits: Universities*. https://www.audit.vic.gov.au/report/results-2017-audits-universities

Victoria University, Melbourne, Australia (2017). *Victoria University's Transformational Agenda* [Internal Unpublished Document]

Vo, H. M., Zhu, C., & Diep, N. A. (2017). The effect of blended learning on student performance at course-level in higher education: A meta-analysis. *Studies in Educational Evaluation, 53*, 17-28. http://doi.org/%2010.1016/j.stueduc.2017.01.002

Vostal, F., & Roberston, S. (2012). Knowledge mediators and lubricating channels: on the temporal politics of remissioning the university. *TOPIA: Canadian Journal of Cultural Studies, 28*, 143-164.

Vytasek, J. M., Patzak, A., & Winne, P. H. (2020). Analytics for Student Engagement. In Virvou M., Alepis E., Tsihrintzis G., Jain L. (eds) *Machine Learning Paradigm. Intelligent Systems Reference Library* (vol 158, pp. 23-48). Springer. https://doi.org/10.1007/978-3-030-13743-4_3

Walker, J. (2009). Time as the fourth dimension in the globalization of higher education. *The Journal of Higher Education, 80*(5), 483-509.

Walsh, J., & Risquez, A. (2020). Using cluster analysis to explore the engagement with a flipped classroom of native and non-native English-speaking management students. *International Journal of Management Education, 18(*2), 1-9. https://doi.org/10.1016/j.ijme.2020.100381

Wlodkowski, R. J. (2003). Accelerated learning in colleges and universities. *New Directions for Adult and Continuing Education, 97*, 5-16.

Zhang, Z. (2016). Multiple imputation with multivariate imputation by chained equation (MICE) package. *Annals of Translational Medicine, 4*(2), 30-35. http://dx.doi.org/10.3978/j.issn.2305-5839.2015.12.63

Zumbach, J., & Funke, J. (2014). Influences of mood on academic course evaluations. *Practical Assessment, Research & Evaluation, 19*(4), 1-12. https://doi.org/10.7275/vbcz-h361

Table 1.
VU block model principles

Design principles	Delivery principles
Have a clear beginning and ending (immersive & self-contained)	Be student-centered, active and engaging (you are the university, 'be fabulous')
Ensure learning outcomes are achievable in the four-week timeframe	Outline the relevance of unit to course and career
Employ a variety of assessment tasks to demonstrate learning outcomes	Provide early and ongoing feedback
Design assessment to be completed within the unit schedule and all feedback returned before commencement of next block	Evaluate students' interests and individual needs/expectations
- Include clear assessment rubrics - Provide opportunities for early student success	Include opportunities for self-assessment that leads to personalised and adaptive learning
Focus on knowledge exploration and application rather than content transmission	Incorporate the use of digital technology
Include opportunities for peer feedback and collaboration	Integrate active and authentic learning practices in all units
Use explicit and differentiated learning opportunities (more than one way to achieve the learning outcome)	
Optimise opportunities to learn in new ways within the parameters of four-week blocks	

Table 2.
Final hypotheses

#	Hypothesis
1	The VU SET and SEU surveys will demonstrate acceptable measurement qualities, validity and reliability, making them worthwhile tools to evaluate the intervention/block model
1a	VU SET and SEU surveys will show acceptable model fit, and discriminant validity in a confirmatory factor analysis
1b	VU SET and SEU surveys will demonstrate content and convergent validity
2	The cohort of first-year students studying in block mode at VU will, overall, be more satisfied and achieve higher grades than prior student cohorts studying in the predecessor model, after accounting for pre-existing trends, covariates and mean structures/nested observations.
2a	The effect of the block on outcomes will be robust to varied sensitivity tests; including across analysis methods and sub-sample choices.
2b	The effect of the block will be moderated by covariates and change in unit assessment design.

Table 3.
Items in the SET and SEU quality surveys at Victoria University

Item #	Student Evaluation of Unit (SEU)	Student Evaluation of Teacher (SET)
1	Overall, I am satisfied with the quality of this unit.	Overall, I am satisfied with the quality of teaching provided by the lecturer / tutor
2	The expectations were clear	This teacher / lecturer gave me helpful feedback
3	The activities helped me to learn	This teacher / lecturer helped make the subject interesting
4	The learning resources were relevant and up to date	This teacher / lecturer made an effort to understand any difficulties I might be having with my work
5	The assessment tasks clearly evaluated the learning outcomes	This teacher / lecturer motivated me to do my best work
6	The workload in this unit was reasonable	This teacher / lecturer was good at explaining things

Note. Anchor points: 1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly agree.

Table 4.
The effect of the block on marks with sensitivity tests

Model	random intercepts included	sample	*n* obs	B(se) [95 CI]
M	unit, student, teacher	responders to SET	15989	6.16(0.58) [5.01, 7.30]
M-S1	unit, student	responders to SET	15989	6.10 (0.53) [5.04, 7.15]
M-S2	unit, student, teacher	responders to SET excl. marks <10	15895	5.91(0.55) [4.82, 6.99]
M-S2^{+}	unit, student	entire sample, missing imputed	86545	11.11(0.34) [10.44, 11.79]
M-S3	unit, student	entire sample, missing imputed, excl. marks <10	varies	9.64(0.28) [9.09, 10.19]

Note. $^{+}$ = Model adopted as the primary result. Model M-S3 sample varies across the 100 imputations as rows mark <10 were deleted across imputations Including a random intercept for teacher listwise deleted any non-responders to the SET.

Table 5.
Significant categorical moderators of the effect of the block on marks

			B (se) [95CI]
Main effect (M-S2) for marks			11.11(0.34) [10.44, 11.79]
Moderator	Interaction term / ANOVA	Sub-group treatment effect	
Attendance type	-4.36(1.04) [-6.39, -2.32]	Full time	11.30(0.35) [10.62, 11.98]
		Part time	6.95(1.05) [4.89, 9.00]
Applicant type	2.24(0.42) [1.41, 3.07]	Direct applicant	10.12(0.39) [9.36, 10.89]
		VTAC applicant	12.37(0.42) [11.55, 13.18]
Gender	1.45(0.41) [0.64, 2.26]	Female	10.46(0.39) [9.69, 11.22]
		Male	11.91(0.41) [11.10, 12.72]
NESB	2.13(0.44) [1.28, 2.99]	ESB	10.36(0.38) [9.61, 11.10]
		NESB	12.49(0.44) [11.62, 13.36]
Discipline	$F(5)=26.46^{***}$	Health	11.00(0.47) [10.09, 11.91]
		Arts and education	7.60(0.53) [6.56, 8.64]
		Business	15.54(0.60) [14.37, 16.71]
		Sport and exercise	11.34(0.62) [10.13, 12.55]
		Law and justice	11.19(0.73) [9.75, 12.63]
		Engineering	11.41(0.85) [9.75, 13.07]
Prior education	$F(4)=4.33^{**}$	Pre secondary school	12.09(0.54) [11.04, 13.14]
		Secondary school	11.80(0.54) [10.75, 12.85]
		Post secondary VET	9.93(0.58) [8.80, 11.07]
		Bachelor	11.27(0.51) [10.27, 12.27]
		Postgraduate	9.97(0.57) [8.85, 11.10]

Note. ESB = English-speaking background. NESB = Non-English speaking background. * $p<.05$, ** $p<.01$, $^{***}p<.001$, where ANOVA = D1 method comparing model including the moderator with main effect model, pooling model results across the 100 imputations.

Table 6.
Effect of the block on satisfaction

	Model A^+ (standardised)	Model A (unstandardized)	Model A-S1 (ordinal, standardised)
Item text	β(se) [95 CI]	B(se) [95 CI]	β (se) [95 CI]
SEU1	-0.14(0.05) [-0.24, -0.04]	-0.12(0.04) [-0.20, -0.04]	-0.31(0.10) [-0.50, -0.11]
SEU2	-0.09(0.05) [-0.18, 0.01]	-0.07(0.04) [-0.16, 0.01]	-0.19(0.10) [-0.38, 0.00]
SEU3	-0.10(0.05) [-0.20, -0.01]	-0.09(0.04) [-0.17, -0.01]	-0.25(0.10) [-0.44, -0.06]
SEU4	-0.16(0.05) [-0.25, -0.06]	-0.13(0.04) [-0.21, -0.05]	-0.34(0.10) [-0.54, -0.15]
SEU5	-0.04(0.05) [-0.13, 0.06]	-0.03(0.04) [-0.10, 0.05]	-0.07(0.10) [-0.27, 0.13]
SEU6	-0.24(0.05) [-0.34, -0.15]	-0.21(0.04) [-0.29, -0.13]	-0.52(0.10) [-0.72, -0.32]
SET1	0.10(0.05) [0.00, 0.19]	0.08(0.04) [0.00, 0.15]	0.13(0.10) [-0.07, 0.33]
SET2	0.13(0.05) [0.03, 0.22]	0.11(0.04) [0.03, 0.19]	0.18(0.10) [-0.01, 0.37]
SET3	0.11(0.05) [0.01, 0.21]	0.10(0.04) [0.01, 0.18]	0.17(0.10) [-0.02, 0.37]
SET4	0.16(0.05) [0.07, 0.25]	0.13(0.04) [0.05, 0.20]	0.28(0.10) [0.09, 0.47]
SET5	0.12(0.05) [0.02, 0.21]	0.10(0.04) [0.02, 0.18]	0.19(0.10) [-0.01, 0.38]
SET6	0.00(0.05) [-0.10, 0.09]	0.00(0.04) [-0.08, 0.07]	-0.08(0.10) [-0.28, 0.12]

Note. $^+$ = Model adopted as the primary result. Model A^+ = Linear, mixed effect regressions with a random intercept for unit, teacher and student, across 100 missing data imputations (*n* observations = 15,989). Model A = Model A^+ standardised. Model A-S1 = Ordinal regression models (cumulative link mixed models) calculated on a single imputed dataset with a random intercept for person, unit and teacher (*n* observations = 15,989).

Table 7.
Significant categorical moderators of the effect of the block on unit satisfaction.

	SEU1 β(se) [95 CI]	SEU2 β(se) [95 CI]	SEU3 β(se) [95 CI]	SEU4 β(se) [95 CI]	SEU5 β(se) [95 CI]	SEU6 β(se) [95 CI]
Main effects	-0.14(0.05) [-0.24, -0.04]	-0.09(0.05) [-0.18, 0.01]	-0.10(0.05) [-0.20, -0.01]	-0.16(0.05) [-0.25, -0.06]	-0.04(0.05) [-0.13, 0.06]	-0.24(0.05) [-0.34, -0.15]
Interaction term or ANOVA						
Applicant type	0.09(0.04) [0.00, 0.17]	*ns*	0.10(0.04) [0.01, 0.18]	0.13(0.04) [0.05, 0.22]	0.12(0.04) [0.03, 0.20]	0.14(0.04) [0.06, 0.23]
Discipline	*ns*	$F(5) = 3.441^{**}$	*ns*	$F(5)=2.933^{*}$	$F(5) = 3.563^{**}$	$F(5) = 6.704^{***}$
Effect of the block across subgroups						
VTAC applicant	-0.19(0.05) [-0.29, -0.08]		-0.15(0.05) [-0.26, -0.05]	-0.22(0.05) [-0.33, -0.12]	-0.10(0.05) [-0.20, 0.01]	-0.32(0.05) [-0.42, -0.21]
Direct applicant [+]	-0.10(0.05) [-0.20, 0.01]		-0.06(0.05) [-0.16, 0.05]	-0.09(0.05) [-0.20, 0.01]	0.02(0.05) [-0.08, 0.13]	-0.17(0.05) [-0.28, -0.07]
Health		-0.13(0.06) [-0.25, -0.01]		-0.22(0.06) [-0.35, -0.10]	-0.06(0.06) [-0.18, 0.06]	-0.35(0.06) [-0.47, -0.23]
Arts and education		0.09(0.07) [-0.04, 0.23]		0.03(0.07) [-0.11, 0.16]	0.13(0.07) [0.00, 0.26]	-0.01(0.07) [-0.14, 0.12]
Business		-0.07(0.08) [-0.24, 0.09]		-0.20(0.08) [-0.37, -0.04]	-0.14(0.08) [-0.30, 0.03]	-0.17(0.08) [-0.33, -0.01]
Sport and exercise		-0.16(0.08) [-0.32, 0.00]		-0.19(0.08) [-0.35, -0.03]	-0.08(0.08) [-0.24, 0.07]	-0.27(0.08) [-0.42, -0.12]
Law and justice		-0.28(0.11) [-0.49, -0.07]		-0.23(0.11) [-0.44, -0.02]	-0.25(0.11) [-0.46, -0.04]	-0.50(0.10) [-0.71, -0.29]
Engineering		-0.11(0.12) [-0.35, 0.12]		-0.17(0.12) [-0.41, 0.07]	0.08(0.12) [-0.15, 0.31]	-0.25(0.12) [-0.47, -0.02]

Note. This table presents significant moderators and marginal effects for the primary model, model A^{+}, table 6. *ns* = non-significant. * p<.05, ** p<.01, ***p<.001, where ANOVA = D1 method comparing model including the moderator with main effect model, pooling model results across

the 100 imputations. Confidence intervals for marginal effects are generated with the Delta method (see Gold, Olin & Wang, 2018). As the scale item was standardised treatment effect reflects average change in standard deviation units.

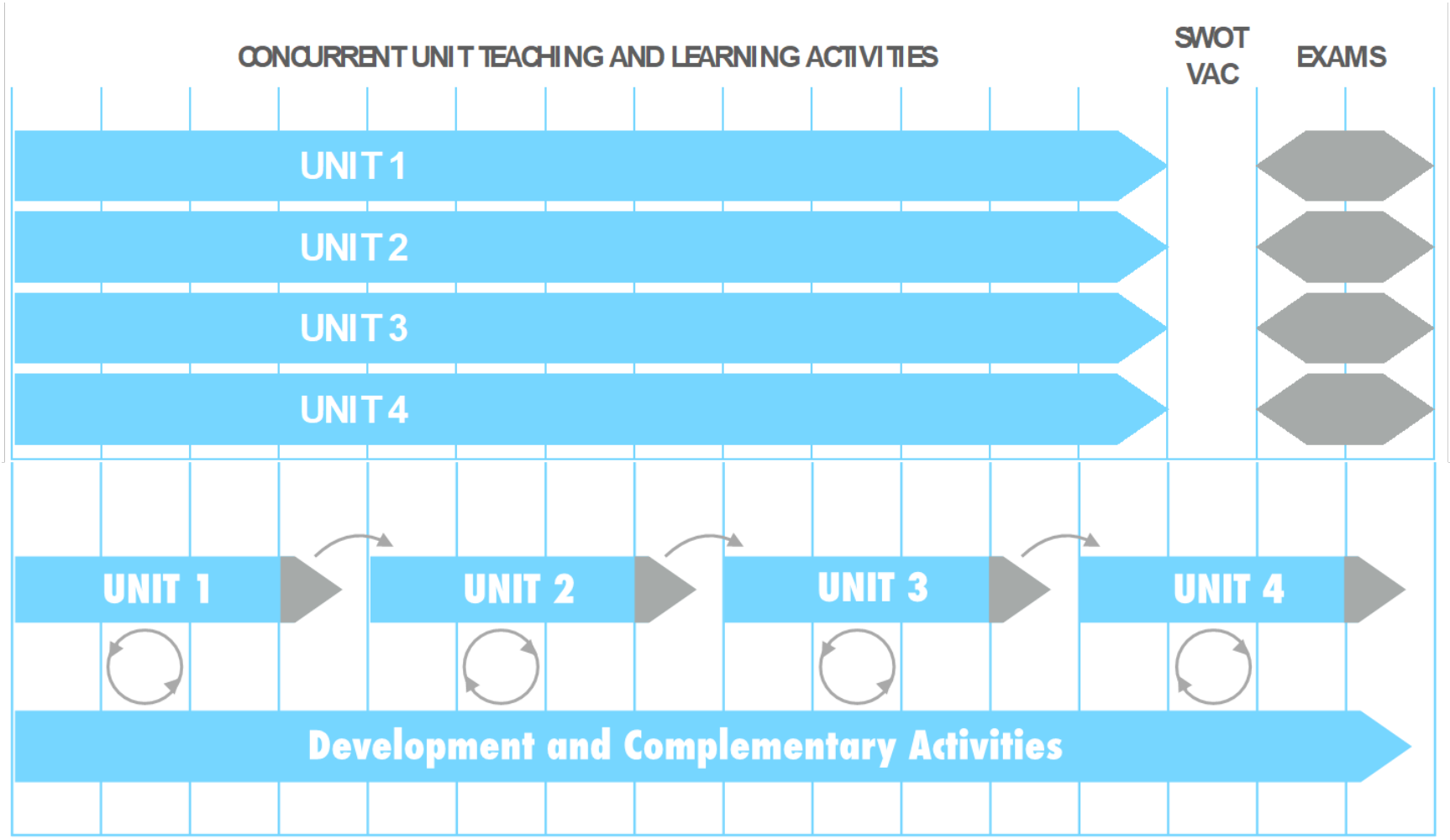
Table 8.
Significant categorical moderators of the effect of the block on teaching satisfaction.

	SET1 β (se) [95CI]	SET2 β (se) [95CI]	SET3 β (se) [95CI]	SET4 β (se) [95CI]	SET5 β (se) [95CI]	SET6 β (se) [95CI]
Main effects	0.10(0.05) [0.00, 0.19]	0.13(0.05) [0.03, 0.22]	0.11(0.05) [0.01, 0.21]	0.16(0.05) [0.07, 0.25]	0.12(0.05) [0.02, 0.21]	0.00(0.05) [-0.10, 0.09]
Interaction term or ANOVA						
Applicant type	0.09(0.04) [0.00, 0.17]	*ns*	*ns*	0.11(0.04) [0.03, 0.20]	0.10(0.04) [0.02, 0.19]	0.08(0.04) [0.00, 0.16]
Discipline	F(5)=2.497*	*ns*	F(5)=2.694*	*ns*	*ns*	*ns*
NESB	*ns*	-0.09(0.05) [-0.19, 0.00]	*ns*	*ns*	*ns*	*ns*
Gender	*ns*	*ns*	0.16(0.04) [0.07, 0.24]	*ns*	*ns*	0.10(0.04) [0.02, 0.19]
Effect of the block across subgroups						
VTAC applicant	0.05(0.05) [-0.05, 0.16]			0.10(0.05) [0.00, 0.20]	0.07(0.05) [-0.04, 0.17]	-0.05(0.05) [-0.15, 0.06]
Direct applicant	0.14(0.05) [0.03, 0.24]			0.21(0.05) [0.11, 0.31]	0.17(0.05) [0.06, 0.27]	0.03(0.05) [-0.07, 0.14]
English primary		0.16(0.05) [0.06, 0.25]				
English secondary		0.06(0.06) [-0.05, 0.18]				
Female			0.05(0.05) [-0.05, 0.15]			-0.04(0.05) [-0.14, 0.06]
Male			0.21(0.06) [0.10, 0.32]			0.06(0.06) [-0.05, 0.17]
Health						
Arts and education	0.14(0.06) [0.01, 0.26]		0.13(0.06) [0.01, 0.25]			
Business	0.20(0.07) [0.06, 0.33]		0.23(0.07) [0.10, 0.37]			
Sport and exercise	0.07(0.08) [-0.09, 0.23]		-0.01(0.08) [-0.17, 0.16]			

Law and justice	-0.01(0.08) [-0.17, 0.15]	0.04(0.08) [-0.11, 0.20]
Engineering	-0.15(0.11) [-0.37, 0.06]	-0.10(0.11) [-0.32, 0.11]

Note. Note. Main effects are from model A^{+} in Table X. *ns* = non-significant. * p<.05, ** p<.01, ***p<.001, where ANOVA = D1 method comparing model including the moderator with main effect model, pooling model results across the 100 imputations. Confidence intervals for marginal effects are generated with the Delta method (see Gold, Olin & Wang, 2018).

Figure 1.
Traditional versus block mode at VU



Note. Study Without Teaching Vacation (SWOTVAC) refers to an exam preparation study period without classes or lectures.

Figure 2.
Block model design and development process

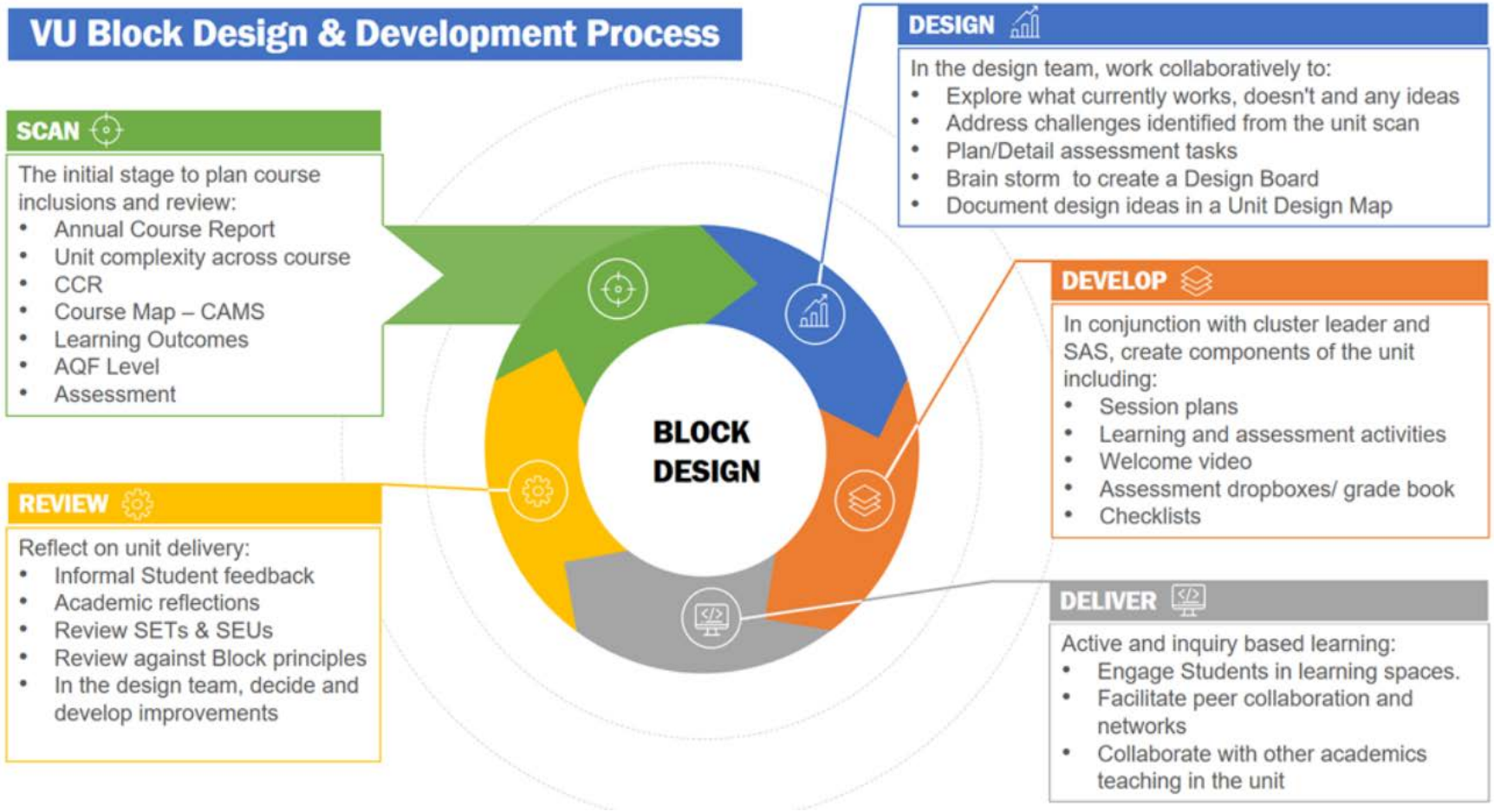
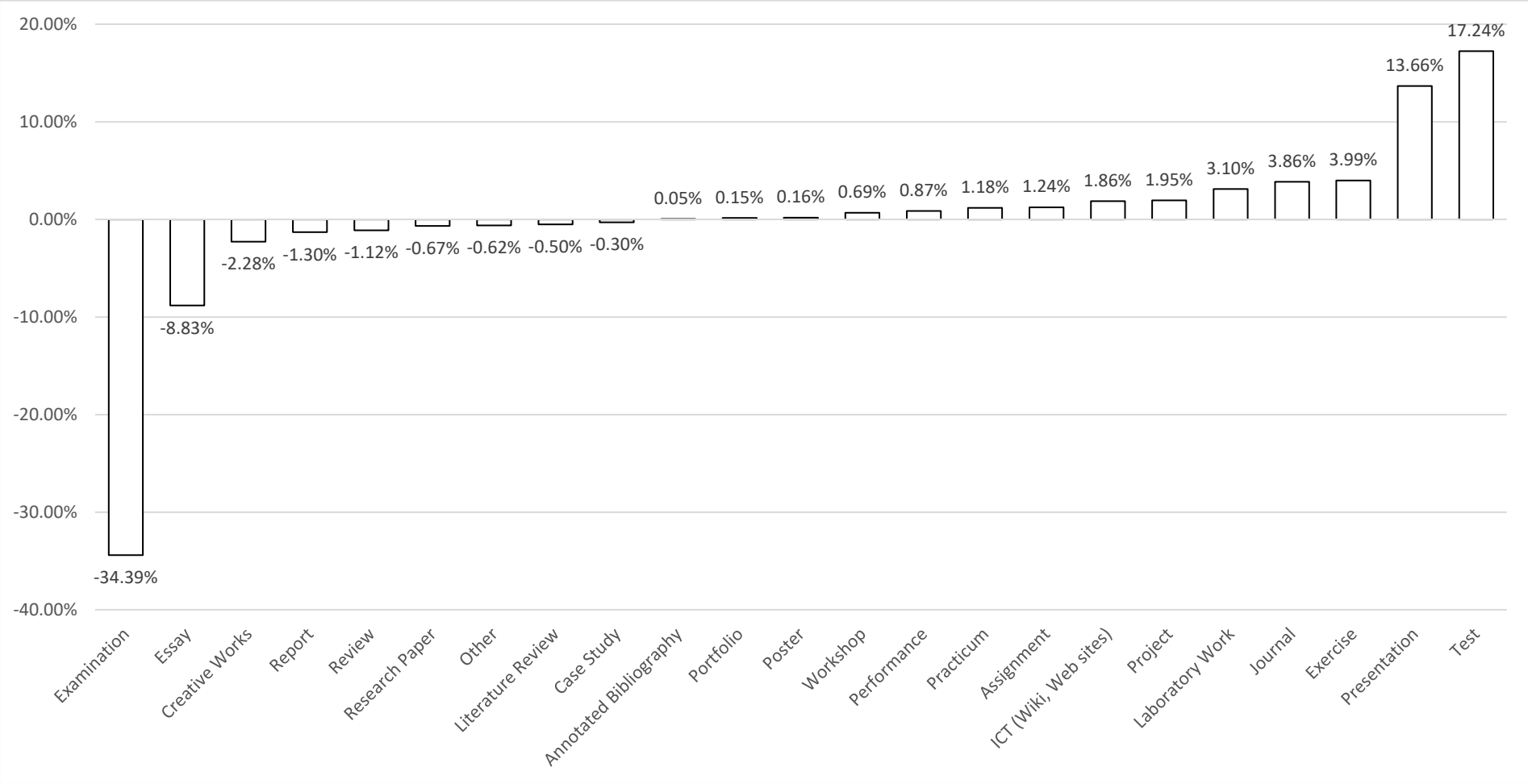


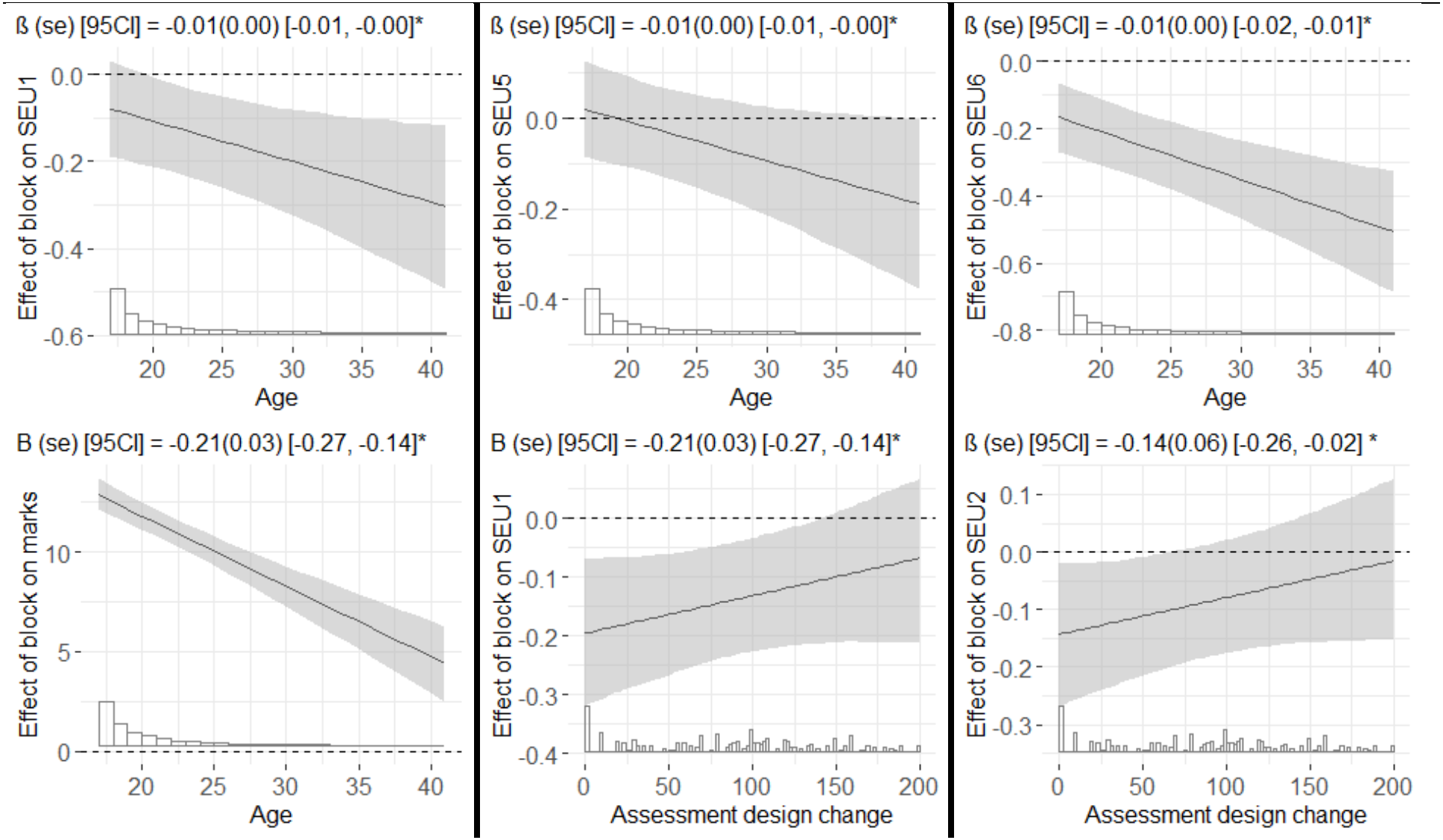
Figure 3.
Change in assessment weightings for assessment types in the transition to block mode, as proportion of total change.



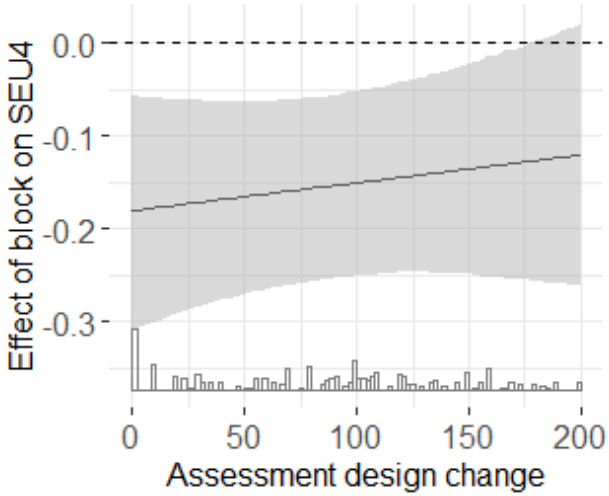
Note. There are no formal definitions for the assessment types in the curriculum management system. Unit Coordinators self-assign a given assessment task to one of the types above during curriculum renewal or development processes. Exams are thought to be more often invigilated

and have a room and time dedicated to their undertaking, whereas tests are thought to be in-class and/or online (email correspondence with curriculum system coordinator, 20/11/2019).

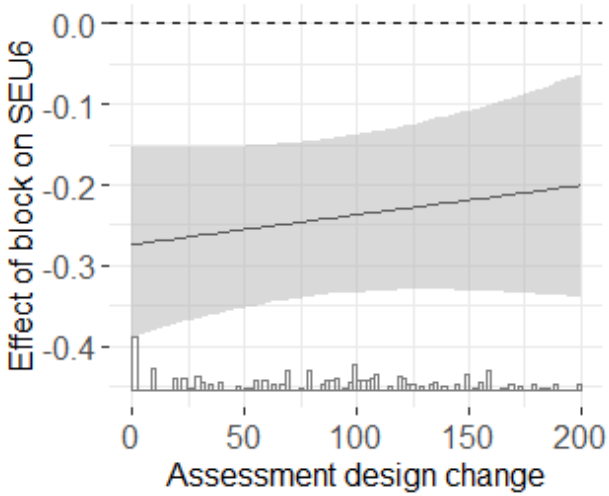
Figure 4.
Significant continuous moderators of the effect of the block.



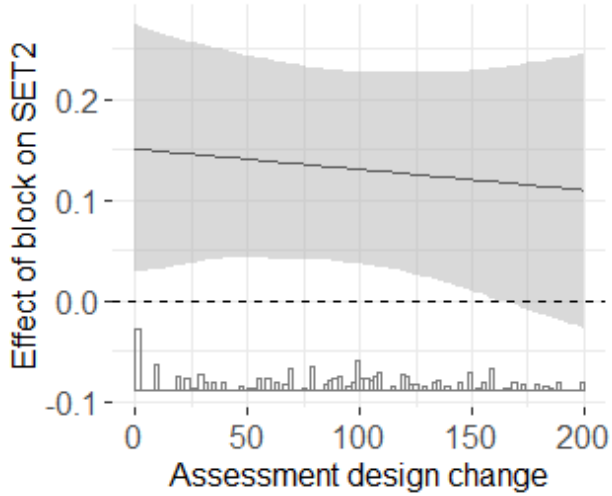
ß (se) [95CI] = -0.18(0.06) [-0.30, -0.06]*



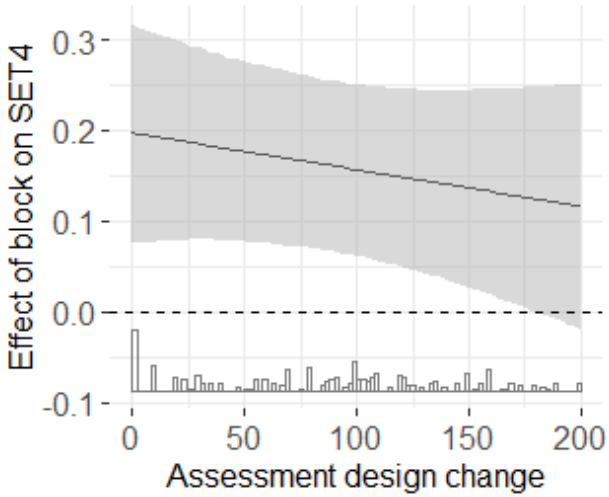
ß (se) [95CI] = -0.27(0.06) [-0.39, -0.15] *



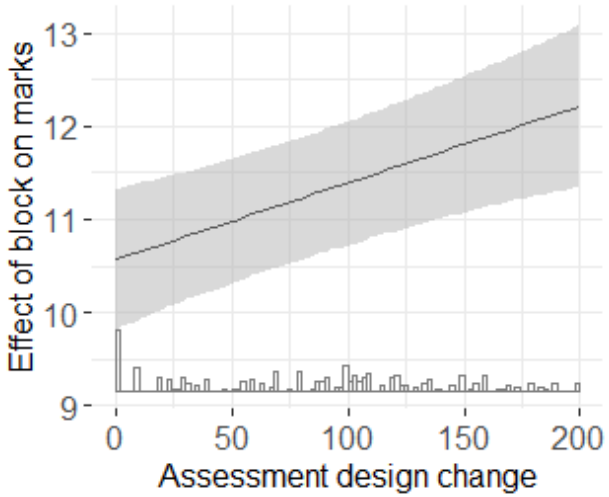
ß (se) [95CI] = -0.14(0.06) [-0.26, -0.02]*



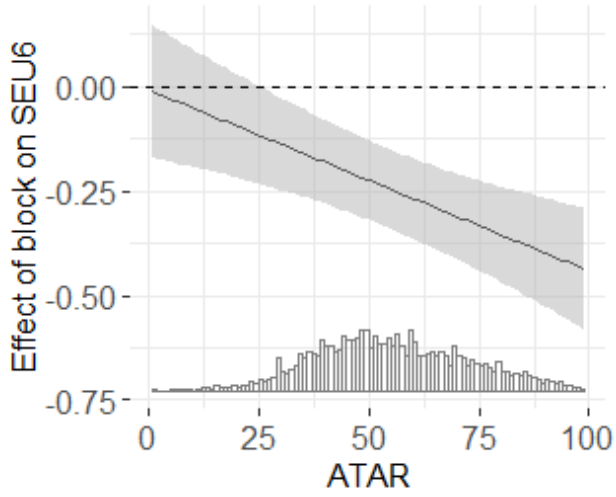
ß (se) [95CI] = 0.20(0.06) [0.08, 0.32]*

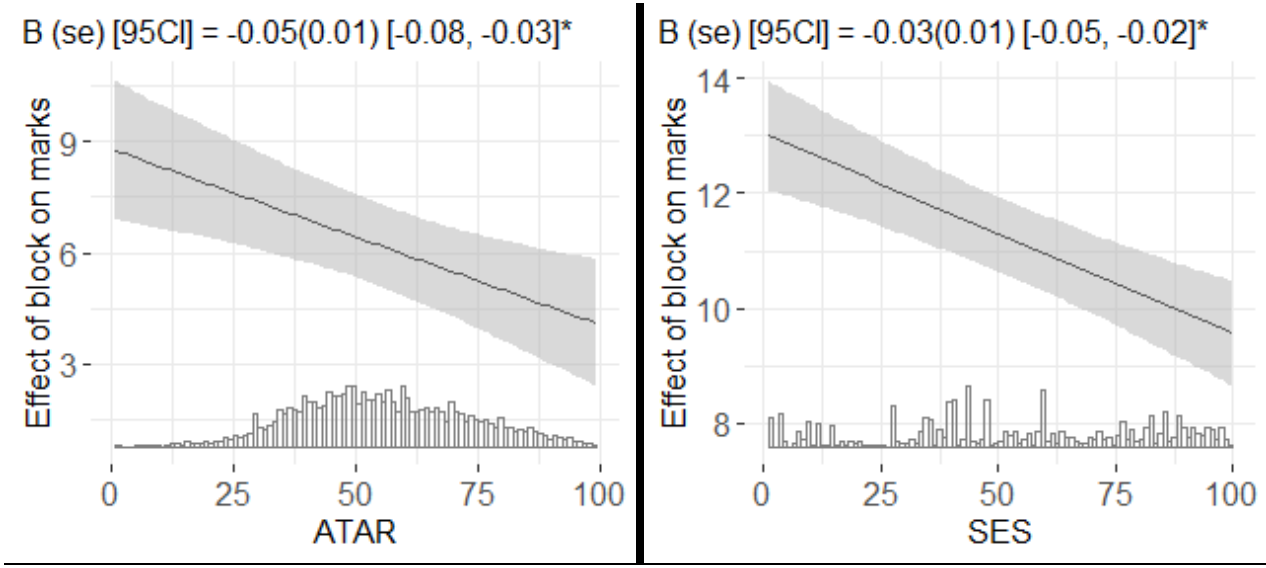


B (se) [95CI] = 0.01(0.00) [0.00, 0.01]*



ß (se) [95CI] = -0.00(0.00) [-0.01, -0.00]*





Note. Shaded area represents the region of significance (see Solt & Hu, 2018). Model specification is per Model A for satisfaction and Model M-S2 for marks.

Introducing block mode to first-year University students: A natural experiment on satisfaction and performance - Supplementary materials

List of sections, tables and figures:

Appendix A: Linear random intercept main effect model results with covariates

As the treatment effect (effect of the block) was the primary interest, all other fixed effect parameters in the main effect models are presented here. In Tables SM-A-1 and 2: VTAC = Victorian Tertiary Admissions Centre, NESB = Non-English Speaking Background, ATSI = Aboriginal or Torres Straight Islander, ATAR = Australian Tertiary Admission Rank, SEIFA = Socio-economic Indexes for Areas, decile, VET = Vocational Education and Training qualification, A&E = Arts and Education, BUS = Business, S&ES = Sport and Exercise Science, L&J = Law and Justice, E&S = Engineering and Science.

Table SM-A-1. Fixed effects in satisfaction models (Model A^{+}).

	SEU1 β (se) [95CI]	SEU2 β (se) [95CI]	SEU3 β (se) [95CI]	SEU4 β (se) [95CI]
Intercept / Constant	0.20(0.10) [0.00, 0.40]	0.16(0.10) [-0.04, 0.36]	0.18(0.10) [-0.02, 0.39]	0.31(0.10) [0.11, 0.52]
Mode - Block	-0.14(0.05) [-0.24, -0.04]	-0.09(0.05) [-0.18, 0.01]	-0.10(0.05) [-0.20, -0.01]	-0.16(0.05) [-0.25, -0.06]
Pathway - Commencing	0.06(0.03) [0.00, 0.11]	0.01(0.03) [-0.04, 0.07]	0.01(0.03) [-0.04, 0.07]	0.01(0.03) [-0.05, 0.06]
Study type - Part time	0.02(0.06) [-0.09, 0.13]	0.02(0.06) [-0.09, 0.13]	0.07(0.06) [-0.04, 0.18]	0.02(0.06) [-0.09, 0.13]
Applicant - VTAC	0.01(0.02) [-0.04, 0.05]	0.00(0.02) [-0.05, 0.05]	-0.02(0.02) [-0.07, 0.03]	-0.02(0.02) [-0.06, 0.03]
Gender - Male	0.05(0.02) [0.01, 0.10]	0.06(0.02) [0.01, 0.11]	0.04(0.02) [-0.01, 0.09]	0.03(0.02) [-0.02, 0.08]
Equity - NESB	0.04(0.03) [-0.02, 0.10]	0.04(0.03) [-0.02, 0.10]	0.06(0.03) [0.00, 0.11]	0.05(0.03) [-0.01, 0.11]
Equity – Birth country	0.08(0.03) [0.02, 0.15]	0.06(0.03) [-0.01, 0.12]	0.05(0.03) [-0.02, 0.11]	0.04(0.03) [-0.02, 0.11]
Equity - ATSI	0.07(0.12) [-0.16, 0.30]	0.08(0.12) [-0.15, 0.31]	0.10(0.12) [-0.13, 0.33]	0.07(0.12) [-0.16, 0.30]
Equity - Disability	-0.09(0.04) [-0.17, -0.01]	-0.12(0.04) [-0.20, -0.04]	-0.11(0.04) [-0.19, -0.03]	-0.08(0.04) [-0.16, 0.00]
Region - Overseas	-0.02(0.08) [-0.17, 0.13]	-0.05(0.08) [-0.20, 0.10]	0.03(0.08) [-0.12, 0.18]	-0.01(0.08) [-0.16, 0.14]
Region - Western	0.03(0.02) [-0.02, 0.07]	0.03(0.02) [-0.01, 0.08]	0.03(0.02) [-0.01, 0.08]	-0.01(0.02) [-0.05, 0.04]
Pathway - International	0.10(0.06) [-0.02, 0.23]	0.20(0.07) [0.07, 0.33]	0.08(0.07) [-0.05, 0.21]	0.00(0.07) [-0.13, 0.12]
Equity - ATAR	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]
Equity - SEIFA	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]
Equity - Age	0.01(0.00) [0.01, 0.02]	0.01(0.00) [0.01, 0.02]	0.01(0.00) [0.01, 0.01]	0.01(0.00) [0.01, 0.01]
Discipline - A&E	0.03(0.05) [-0.07, 0.14]	0.02(0.05) [-0.08, 0.12]	0.05(0.05) [-0.05, 0.15]	0.04(0.05) [-0.06, 0.15]
Discipline - BUS	0.07(0.07) [-0.06, 0.20]	0.02(0.07) [-0.11, 0.16]	0.04(0.07) [-0.08, 0.17]	0.05(0.07) [-0.08, 0.18]
Discipline – S&ES	0.09(0.06) [-0.03, 0.21]	0.04(0.06) [-0.08, 0.16]	0.12(0.06) [0.00, 0.23]	0.14(0.06) [0.02, 0.26]
Discipline - L&J	0.12(0.08) [-0.03, 0.28]	0.12(0.08) [-0.03, 0.27]	0.15(0.07) [0.00, 0.30]	0.14(0.08) [-0.01, 0.29]
Discipline - E&S	-0.05(0.09) [-0.22, 0.12]	-0.06(0.08) [-0.23, 0.10]	-0.08(0.08) [-0.24, 0.08]	-0.08(0.08) [-0.25, 0.08]
Multimodal	0.15(0.06) [0.04, 0.26]	0.11(0.06) [0.00, 0.23]	0.14(0.06) [0.03, 0.25]	0.20(0.06) [0.09, 0.32]
Education - Bachelor	0.02(0.04) [-0.04, 0.09]	0.04(0.04) [-0.03, 0.11]	0.02(0.04) [-0.05, 0.09]	0.05(0.04) [-0.02, 0.12]
Education - VET	0.01(0.04) [-0.06, 0.09]	0.03(0.04) [-0.04, 0.11]	0.00(0.04) [-0.08, 0.07]	0.02(0.04) [-0.06, 0.09]
Education - Year 12	-0.01(0.04) [-0.09, 0.06]	-0.03(0.04) [-0.10, 0.05]	-0.03(0.04) [-0.10, 0.04]	-0.02(0.04) [-0.09, 0.05]
Education - sub Year 12	0.02(0.04) [-0.06, 0.09]	0.00(0.04) [-0.08, 0.07]	0.03(0.04) [-0.05, 0.10]	0.04(0.04) [-0.04, 0.11]
Linear time trend	0.01(0.01) [-0.01, 0.03]	0.01(0.01) [-0.01, 0.02]	0.01(0.01) [0.00, 0.03]	0.00(0.01) [-0.01, 0.02]
Quadratic time trend	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]

Table SM-A-1. Continued.

	SEU5 β (se) [95CI]	SEU6 β (se) [95CI]	SET1 β (se) [95CI]	SET2 β (se) [95CI]
Intercept / Constant	0.20(0.10) [0.00, 0.40]	0.29(0.10) [0.08, 0.49]	0.06(0.10) [-0.13, 0.26]	0.16(0.10) [-0.04, 0.35]
Mode - Block	-0.04(0.05) [-0.13, 0.06]	-0.24(0.05) [-0.34, -0.15]	0.10(0.05) [0.00, 0.19]	0.13(0.05) [0.03, 0.22]
Pathway - Commencing	0.04(0.03) [-0.02, 0.10]	0.00(0.03) [-0.06, 0.05]	-0.03(0.03) [-0.08, 0.03]	-0.06(0.03) [-0.11, 0.00]
Study type - Part time	0.03(0.06) [-0.08, 0.15]	0.03(0.06) [-0.08, 0.14]	0.03(0.05) [-0.08, 0.13]	0.01(0.05) [-0.09, 0.12]
Applicant - VTAC	-0.01(0.02) [-0.06, 0.04]	-0.01(0.02) [-0.06, 0.04]	-0.02(0.02) [-0.07, 0.03]	-0.06(0.02) [-0.11, -0.02]
Gender - Male	0.03(0.02) [-0.02, 0.08]	0.08(0.02) [0.03, 0.12]	0.08(0.02) [0.03, 0.12]	0.09(0.02) [0.04, 0.13]
Equity - NESB	0.05(0.03) [-0.01, 0.11]	0.00(0.03) [-0.06, 0.06]	0.01(0.03) [-0.05, 0.07]	0.02(0.03) [-0.03, 0.08]
Equity – Birth country	0.02(0.03) [-0.05, 0.08]	0.00(0.03) [-0.07, 0.06]	0.01(0.03) [-0.05, 0.07]	0.03(0.03) [-0.03, 0.09]
Equity - ATSI	0.05(0.12) [-0.18, 0.28]	0.08(0.12) [-0.15, 0.31]	0.12(0.11) [-0.10, 0.34]	0.05(0.11) [-0.17, 0.28]
Equity - Disability	-0.08(0.04) [-0.17, 0.00]	-0.16(0.04) [-0.24, -0.09]	-0.09(0.04) [-0.17, -0.02]	-0.13(0.04) [-0.20, -0.05]
Region - Overseas	0.02(0.08) [-0.13, 0.17]	-0.05(0.07) [-0.20, 0.10]	-0.02(0.07) [-0.16, 0.12]	0.01(0.07) [-0.13, 0.15]
Region - Western	0.01(0.02) [-0.03, 0.06]	0.00(0.02) [-0.05, 0.04]	0.04(0.02) [-0.01, 0.08]	0.01(0.02) [-0.04, 0.05]
Pathway - International	0.09(0.07) [-0.04, 0.22]	0.03(0.06) [-0.10, 0.16]	0.08(0.06) [-0.04, 0.21]	0.10(0.06) [-0.03, 0.22]
Equity - ATAR	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]
Equity - SEIFA	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]
Equity - Age	0.01(0.00) [0.01, 0.01]	0.00(0.00) [0.00, 0.01]	0.01(0.00) [0.01, 0.02]	0.01(0.00) [0.01, 0.02]
Discipline - A&E	0.06(0.05) [-0.04, 0.16]	0.02(0.05) [-0.08, 0.13]	-0.03(0.05) [-0.13, 0.07]	0.02(0.05) [-0.08, 0.11]
Discipline - BUS	-0.01(0.07) [-0.14, 0.12]	0.11(0.07) [-0.03, 0.24]	0.00(0.06) [-0.12, 0.13]	0.03(0.06) [-0.09, 0.16]
Discipline – S&ES	0.06(0.06) [-0.06, 0.18]	0.11(0.06) [-0.01, 0.24]	-0.02(0.06) [-0.13, 0.09]	0.08(0.06) [-0.03, 0.19]
Discipline - L&J	0.13(0.08) [-0.01, 0.28]	0.23(0.08) [0.08, 0.38]	0.06(0.07) [-0.09, 0.20]	0.02(0.07) [-0.12, 0.17]
Discipline - E&S	-0.15(0.08) [-0.31, 0.01]	0.00(0.09) [-0.17, 0.17]	-0.16(0.08) [-0.32, 0.01]	-0.10(0.08) [-0.26, 0.05]
Multimodal	0.22(0.06) [0.10, 0.33]	0.01(0.06) [-0.10, 0.13]	0.15(0.06) [0.04, 0.26]	0.17(0.06) [0.06, 0.28]
Education - Bachelor	0.03(0.04) [-0.04, 0.10]	0.02(0.04) [-0.05, 0.09]	0.03(0.03) [-0.04, 0.09]	0.02(0.03) [-0.05, 0.08]
Education - VET	0.03(0.04) [-0.05, 0.11]	0.00(0.04) [-0.07, 0.08]	0.04(0.04) [-0.04, 0.11]	0.01(0.04) [-0.06, 0.09]
Education - Year 12	0.00(0.04) [-0.08, 0.07]	-0.01(0.04) [-0.08, 0.07]	0.01(0.04) [-0.06, 0.08]	0.00(0.04) [-0.07, 0.07]
Education - sub Year 12	0.02(0.04) [-0.06, 0.10]	0.03(0.04) [-0.05, 0.10]	0.02(0.04) [-0.05, 0.09]	0.00(0.04) [-0.07, 0.08]
Linear time trend	0.01(0.01) [0.00, 0.03]	0.01(0.01) [-0.01, 0.02]	0.03(0.01) [0.01, 0.05]	0.02(0.01) [0.00, 0.04]
Quadratic time trend	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]

Table SM-A-1. Continued

	SET3 β (se) [95CI]	SET4 β (se) [95CI]	SET5 β (se) [95CI]	SET6 β (se) [95CI]
Intercept / Constant	0.03(0.10) [-0.16, 0.23]	0.12(0.10) [-0.07, 0.32]	0.03(0.10) [-0.17, 0.23]	0.13(0.10) [-0.07, 0.32]
Mode - Block	0.11(0.05) [0.01, 0.21]	0.16(0.05) [0.07, 0.25]	0.12(0.05) [0.02, 0.21]	0.00(0.05) [-0.10, 0.09]
Pathway - Commencing	-0.03(0.03) [-0.09, 0.02]	-0.05(0.03) [-0.11, 0.01]	-0.04(0.03) [-0.10, 0.02]	-0.05(0.03) [-0.10, 0.01]
Study type - Part time	0.07(0.05) [-0.04, 0.17]	-0.03(0.05) [-0.14, 0.08]	0.07(0.06) [-0.04, 0.18]	0.06(0.05) [-0.04, 0.17]
Applicant - VTAC	-0.03(0.02) [-0.08, 0.01]	-0.07(0.02) [-0.12, -0.03]	-0.05(0.02) [-0.09, 0.00]	-0.04(0.02) [-0.09, 0.01]
Gender - Male	0.05(0.02) [0.01, 0.10]	0.07(0.02) [0.03, 0.12]	0.01(0.02) [-0.03, 0.06]	0.07(0.02) [0.02, 0.11]
Equity - NESB	-0.01(0.03) [-0.07, 0.05]	0.05(0.03) [0.00, 0.11]	0.02(0.03) [-0.04, 0.08]	0.02(0.03) [-0.04, 0.07]
Equity – Birth country	0.03(0.03) [-0.03, 0.10]	0.00(0.03) [-0.07, 0.06]	0.05(0.03) [-0.01, 0.12]	0.04(0.03) [-0.02, 0.10]
Equity - ATSI	0.03(0.11) [-0.19, 0.25]	0.12(0.12) [-0.11, 0.35]	0.03(0.12) [-0.20, 0.26]	0.06(0.11) [-0.16, 0.28]
Equity - Disability	-0.08(0.04) [-0.15, 0.00]	-0.12(0.04) [-0.20, -0.04]	-0.11(0.04) [-0.19, -0.03]	-0.06(0.04) [-0.13, 0.02]
Region - Overseas	-0.01(0.07) [-0.15, 0.13]	0.02(0.07) [-0.12, 0.17]	0.03(0.08) [-0.11, 0.18]	-0.04(0.07) [-0.18, 0.10]
Region - Western	0.02(0.02) [-0.02, 0.06]	0.03(0.02) [-0.02, 0.07]	0.01(0.02) [-0.04, 0.06]	0.03(0.02) [-0.01, 0.08]
Pathway - International	0.08(0.06) [-0.04, 0.20]	0.07(0.06) [-0.05, 0.20]	0.06(0.06) [-0.07, 0.18]	0.11(0.06) [-0.01, 0.23]
Equity - ATAR	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]
Equity - SEIFA	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]
Equity - Age	0.01(0.00) [0.01, 0.02]	0.01(0.00) [0.01, 0.01]	0.01(0.00) [0.01, 0.02]	0.01(0.00) [0.01, 0.01]
Discipline - A&E	0.02(0.05) [-0.08, 0.12]	0.01(0.05) [-0.09, 0.10]	-0.01(0.05) [-0.11, 0.08]	0.02(0.05) [-0.08, 0.11]
Discipline - BUS	0.01(0.06) [-0.12, 0.13]	0.00(0.06) [-0.12, 0.12]	0.02(0.06) [-0.11, 0.14]	0.01(0.06) [-0.12, 0.13]
Discipline – S&ES	0.06(0.06) [-0.06, 0.17]	0.04(0.06) [-0.07, 0.15]	0.05(0.06) [-0.06, 0.17]	0.02(0.06) [-0.09, 0.14]
Discipline - L&J	0.11(0.07) [-0.04, 0.25]	0.06(0.07) [-0.08, 0.20]	0.08(0.07) [-0.06, 0.23]	0.09(0.07) [-0.06, 0.23]
Discipline - E&S	-0.09(0.08) [-0.25, 0.08]	-0.09(0.08) [-0.24, 0.07]	-0.05(0.08) [-0.21, 0.11]	-0.14(0.08) [-0.30, 0.02]
Multimodal	0.15(0.06) [0.04, 0.26]	0.12(0.06) [0.00, 0.23]	0.18(0.06) [0.07, 0.30]	0.11(0.06) [0.00, 0.22]
Education - Bachelor	0.01(0.03) [-0.05, 0.08]	0.01(0.04) [-0.06, 0.08]	0.00(0.04) [-0.07, 0.07]	0.03(0.03) [-0.03, 0.10]
Education - VET	0.00(0.04) [-0.07, 0.08]	0.03(0.04) [-0.05, 0.10]	0.01(0.04) [-0.07, 0.09]	0.04(0.04) [-0.03, 0.11]
Education - Year 12	-0.02(0.04) [-0.09, 0.05]	-0.02(0.04) [-0.09, 0.05]	-0.02(0.04) [-0.10, 0.05]	0.03(0.04) [-0.04, 0.10]
Education - sub Year 12	-0.01(0.04) [-0.08, 0.07]	0.01(0.04) [-0.07, 0.08]	0.00(0.04) [-0.08, 0.07]	0.04(0.04) [-0.03, 0.11]
Linear time trend	0.02(0.01) [0.00, 0.03]	0.02(0.01) [0.00, 0.04]	0.02(0.01) [0.00, 0.04]	0.00(0.01) [-0.01, 0.02]
Quadratic time trend	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]	0.00(0.00) [0.00, 0.00]

Table A-SM-2. Fixed effects in marks model (Model M-S2^{+}).

	Marks B (se) [95CI]
Intercept / Constant	41.35(1.23) [38.94, 43.76]
Mode - Block	11.11(0.34) [10.44, 11.79]
Pathway - Commencing	-0.64(0.22) [-1.07, -0.21]
Study type - Part time	1.01(0.71) [-0.38, 2.41]
Applicant - VTAC	0.08(0.29) [-0.50, 0.65]
Gender - Male	-3.76(0.32) [-4.39, -3.12]
Equity - NESB	-3.15(0.41) [-3.95, -2.35]
Equity – Birth country	-1.38(0.44) [-2.24, -0.51]
Equity - ATSI	-5.14(1.75) [-8.57, -1.70]
Equity - Disability	-2.21(0.62) [-3.43, -0.99]
Region - Overseas	4.41(1.12) [2.22, 6.60]
Region - Western	0.22(0.33) [-0.42, 0.86]
Pathway - International	5.15(0.92) [3.35, 6.95]
Equity - ATAR	0.21(0.01) [0.19, 0.23]
Equity - SEIFA	0.03(0.01) [0.01, 0.04]
Equity - Age	0.28(0.03) [0.23, 0.33]
Discipline - A&E	-1.66(0.60) [-2.85, -0.48]
Discipline - BUS	-0.98(0.72) [-2.38, 0.43]
Discipline – S&ES	0.67(0.68) [-0.66, 2.00]
Discipline - L&J	2.67(0.79) [1.12, 4.22]
Discipline - E&S	0.63(0.92) [-1.17, 2.42]
Multimodal	-0.57(0.45) [-1.44, 0.30]
Education - Bachelor	0.04(0.45) [-0.84, 0.93]
Education - VET	0.26(0.48) [-0.68, 1.20]
Education - Year 12	-0.74(0.47) [-1.65, 0.17]
Education - sub Year 12	0.20(0.47) [-0.72, 1.12]
Linear time trend	0.33(0.05) [0.22, 0.44]
Quadratic time trend	-0.02(0.00) [-0.03, -0.02]

Table A-SM-3. Steps applied to generate current sample

Filter records to: Reportable, bachelor, domestic and 'international Melbourne' student records.
Filter records to only include units with enrolments and at least 1 response to the SET survey pre block (meaning the core semester 1 and 2 delivery periods in 2016 or 2017), and post block mode introduction (meaning the core 8 delivery periods in 2018). Treatment effect: Recode delivery periods so that traditional = 1 and block = 2. Delivery period is a completely reliable indicator of block or traditional delivery. All unit codes began with a numerical suffix of '1' in the present study. Note the numerical suffix may not be a reliable indicator of the year-level of a given unit in all cases. Time variable: Each Semester period = 4, each block period = 1, beginning in Semester 1 2016. For example, 2016 Semester 1 = 4, 2016 Semester =8…2018 Block 1 = 17, 2018 Semester 1 = 20

Note. This information is primarily intended for future analysts at the intervention University, in order to replicate the current sample.

Appendix B: Content, convergent and psychometric validity for the VU SET and SEU measure of student satisfaction.

Content validity. The items in the scale were developed by a panel of University leadership, administrators and academics in 2016, with an earlier version of the scales pre-dating those used in the present study. The content of the items align well with many SET-type surveys, and the Australian national Quality Indicators of Learning and Teaching survey (QILT). For example, aligned SES items tap into course work being "well-structured and focused" (item "STDSTRUC", see SEU2 and SEU3), "relevant to your education as a whole" (STDRELEV, see SEU4); and teachers demonstrating "concern for student learning" (TCHCONLR, see SET4), and "provide clear explanations on coursework and assessment" (TCHCLEXP, see SET6 and SEU2), and "commented on your work in ways that help you learn" (TCHFEEDB, see SET2).

Similarly, content of the SET and SEU items align with domains of more validated student evaluation of teaching scales. Helpfully, a summary table of domains measured in SET-like scales is presented by Spooren et al. (2013, see Table 1, p. 606 - 607). There is clear alignment between the SET and SEU items and, for example, the dimensions of course rigor, quality of instruction, instructor helpfulness in the institutional quality tool used in Barth (2008); the appropriate assessment and workload in the Ginns et al (2007) SCEQ scale; and organisation/clarity, workload difficult, individual rapport, and other dimensions of Marsh and colleagues' (1982) SEEQ scale.

Convergent validity. We test convergent validity by associations with student marks and the national Australian Government tertiary student satisfaction survey (QILT, 2018). Student marks for the associated unit delivery period were significantly and positively predicted by both the SET and SEU latent factors. With each increase in the latent factor score for SEU, marks

increased (β = 1.064 [0.595, 1.533]), as with each increase in the SET latent score (β = 1.948 [1.471, 2.424]).

Table SM-B-1. Zero-order correlation matrix for SET and SEU items

Variable		1	2	3	4	5	6	7	8	9	10	11	12	13
1	SEU1	-												
2	SEU2	0.77	-											
3	SEU3	0.72	0.71	-										
4	SEU4	0.76	0.72	0.76	-									
5	SEU5	0.70	0.72	0.72	0.68	-								
6	SEU6	0.61	0.56	0.60	0.58	0.55	-							
7	SET1	0.68	0.63	0.59	0.62	0.57	0.44	-						
8	SET2	0.61	0.59	0.57	0.59	0.54	0.43	0.79	-					
9	SET3	0.68	0.65	0.57	0.59	0.56	0.45	0.82	0.75	-				
10	SET4	0.59	0.58	0.55	0.57	0.53	0.42	0.79	0.82	0.73	-			
11	SET5	0.66	0.64	0.59	0.62	0.57	0.45	0.82	0.79	0.84	0.79	-		
12	SET6	0.65	0.61	0.58	0.61	0.56	0.44	0.86	0.76	0.81	0.76	0.78	-	
13	Mark	0.12	0.13	0.13	0.11	0.12	0.14	0.14	0.14	0.15	0.15	0.14	0.13	-

Note. Pearson correlations were calculated and all are significant at $p<.001$, even after Holm-Bonferroni correction for 144 multiple comparisons.

We obtained a matched sub-sample of students who completed both the SET and SEU survey, as well as the Australian Government Department of Education-funded national student satisfaction survey. In the sample of first-year students at VU from years 2016-2018, we were successfully able to identify between 1016 and 2583 students who completed a SET and SEU survey during the same year as the SES. As the SES items ask students to rate their overall experience with the University that year, SET and SEU scores for all units in the same year were aggregated. We chose 6 derived variables and one item from the SES that have some alignment with the SET and SEU. Readers are directed to the 2018 SES data dictionary for specifics on the scale items and constructs (Social Research Centre, 2018). Pearsons correlation coefficients are reported in Table SM-B-2. We find small to moderate relationships between the twelve SET and SEU items and the SES constructs selected, with teaching-related items more consistently correlated.

Table SM-B-2. Zero-order correlation matrix for SET, SEU, marks and SES constructs

	ENGAGE	OVERALL	QLTEACH	TEACH	RESOURCE	SUPPORT
SEU1	0.18	0.29	0.32	0.36	0.27	0.27
SEU2	0.19	0.27	0.31	0.36	0.28	0.28
SEU3	0.20	0.28	0.30	0.37	0.28	0.30
SEU4	0.17	0.26	0.29	0.35	0.27	0.27
SEU5	0.18	0.27	0.30	0.35	0.29	0.29
SEU5	0.17	0.25	0.26	0.31	0.23	0.27
SEU6	0.15	0.22	0.32	0.33	0.22	0.22
SET1	0.14	0.21	0.31	0.33	0.21	0.23
SET2	0.17	0.22	0.30	0.35	0.23	0.24
SET3	0.17	0.21	0.31	0.35	0.24	0.25
SET4	0.15	0.21	0.30	0.33	0.23	0.23
SET5	0.17	0.24	0.32	0.36	0.25	0.25
SET6	0.18	0.29	0.32	0.36	0.27	0.27

Note. Pearson correlations were calculated, and all are significant at $p<001$. Sample sizes vary for each derived variable due to missing data in the SES, and are as follows: Develop (n=2517). Engage (n=2582). Overall (n=2582). Resource (n=2486). Support (n=2222), and Teach (n=2559). Complete definitions and details on the metrics produced by the SES can be found in the SES data dictionary 2018 (available on request from the Social Research Centre).

Internal structure. The correlation structure of SET, SEU and marks are in Table SM-B-1. Hu and Bentler (1999) suggest benchmarks of Comparative Fix Index (CFI) ≥ 0.95, Root Mean Square Error of Approximation (RMSEA) ≤ 0.06, and Standardized Root Mean Square Residual (SRMR) ≤ 0.08 to describe a good model fit. The CFA results of a model with the SEU and SET items loading on two co-varied latent factors provides a good fit, with exception of the RMSEA, although this fit statistic is biased against simple models. With n=15,982 observations, standardised item loadings were high (SEU 0.742 - 0.944; SET 0.944 - 1.067), and fit results were as follows: $\chi^2(53)=6055.17$, p<.000, CFI = 0.969, TLI = 0.961, RMSEA = 0.084 [90% CI 0.082, 0.086], SRMR = 0.022. Standardised covariance between the SET and SEU factors was significant and large ($\Psi = 0.789$ [0.782, 0.796]). The factors were discriminant, however. In line with the discriminant validity test in Bagozzi et al. (1991), when a model was estimated with the

12 SET and SEU indicators loading on a single factor, model fit significantly worsened ($\Delta\chi^2(1) = 25395$, $p<.001$; model fit for single factor solution $\chi^2(54) = 32805.68$, $p<.000$; CFI = 0.84, TLI = 0.79, RMSEA = 0.18 [90% CI 0.18, 0.19], SRMR = 0.07; standardised factor loadings ranged from 0.68 – 1.11), supporting a two factor solution. Student marks for the associated unit delivery period were significantly and positively predicted by both the SET and SEU latent factors. With each increase in the latent factor score for SEU, marks increased (β = 1.064 [0.595, 1.533]), as with each increase in the SET latent score (β = 1.948 [1.471, 2.424]).

Together, these preliminary analyses support the value of examining the effect of introducing block on the SET and SEU items. While the latent correlation is high between items that, in content, target characteristics of a unit and characteristics of teaching, a test of discriminant validity supported a two-factor solution. The content of the items aligns with other established measures of student satisfaction with teaching; and importantly, the content also aligns with the principles of pedagogy and unit design articulated in the block model intervention. Relationships with the SES survey were mostly small, with exception of teaching-related items which tended to be more moderately sized. The negative skew of the item scores however suggests likely range restriction.

Appendix C: Further detail on Australian Higher Education for international readers.

For the benefit of international readers this section introduces the Australian University sector and gives national context to the University in focus. Australian universities are relatively young with the first PhD thought to be awarded in 1948 (Dobson, 2012), and were modeled on UK and Scottish systems (Gascoigne, 1996). There are now 40 Universities in Australia, each created through an Act of Parliament (State), catering for 1,082,533 students in 2018 (Department of Education and Training, 2019). The sector is characterised by historical substantial public investment and regulation, but with a growing reliance on international student fees and philanthropy. Based on 2016 data, the Organisation for Economic Co-operation and Development [OECD] (2019) *Education at a Glance* report indicates Australia invests 3.6% of its total Government expenditure on tertiary education (see Indicator C4.1), ranking Australia tenth in the world by that metric (Figure SM-C-1). Tertiary Education Quality and Standards Agency (TEQSA) regulates the sector, although Universities are self-accrediting. A key requirement to retain the status of a University in Australia, which is uniquely privileged among higher education provider types, is the production of high quality research, reaffirmed in a recent commissioned review (Coaldrake, 2019).

Contemporary trends influencing Australian Universities are mirrored globally. In line with the then Australian Labor Government goal of increasing access and attainment of University education, Government caps on the number of places offered at Universities were lifted for a time, and access increased (Bradley et al., 2008, Dow, 2013). Australia was one of the earliest developed nations to increase access, and although a subsequent Liberal Australian Government later reinstated caps on placements, now has the highest tertiary admittance/entrance rate of all OECD nations (OECD, 2019). With the transition from elite to

mass in Australia (Bradley et al., 2008), understanding and supporting student retention, satisfaction and success became correspondingly more important. The Federal Government introduced a student satisfaction ranking system some time ago, with rankings published to assist students (QILT, 2017), and there are proposals to make University funding contingent on pass rates, retention rates and satisfaction.

In Australia, there is some positive evidence of successful expansion, but also areas for further improvement. A twenty year non-matched cohort study found improvement in first-year University student experiences, including a smoother transition from secondary school, albeit in samples from Australia's more research-intensive and elite institutions (Baik et al., 2015). By contrast, the national student experience survey indicates little change in satisfaction from 2012-2018, with the exception of increased student support from 2013 onwards (QILT, 2017, Figure 1). A more in-depth analysis of students with 'low' tertiary entrance rank scores, defined as lower than the 70th percentile of the ATAR score, in the aforementioned panel study found these students scored differently from the overall trends in the time series. These students were more like to be from regional areas, first in family to attend University, report having a disability, and be less prepared, academically engaged and at greater risk of attrition than students with higher ATARs (Baik et al., 2019); and also more stressed (Naylor et al., 2018).

Similarly, sector-wide attrition rates have remained stable over a ten-year period at around 15% (Higher Education Standards Panel, 2017), which given the expansion of the sector can be considered a positive result. The lack of further improvement in retention may be due in part to many causes of attrition being primarily personal, unrelated to studies and beyond the control of the University (Harvey et al., 2017). Harvey and colleagues also tracked students over time using a unique identifier, and found the majority of attritions re-entered the system within

the next seven years, perhaps more in line with changing career goals or interests. Nonetheless, overcoming inequity in student success remains key to the sector in Australia, and it is constitutive of the mission and values of the focal University.

Victoria University, Melbourne, Australia

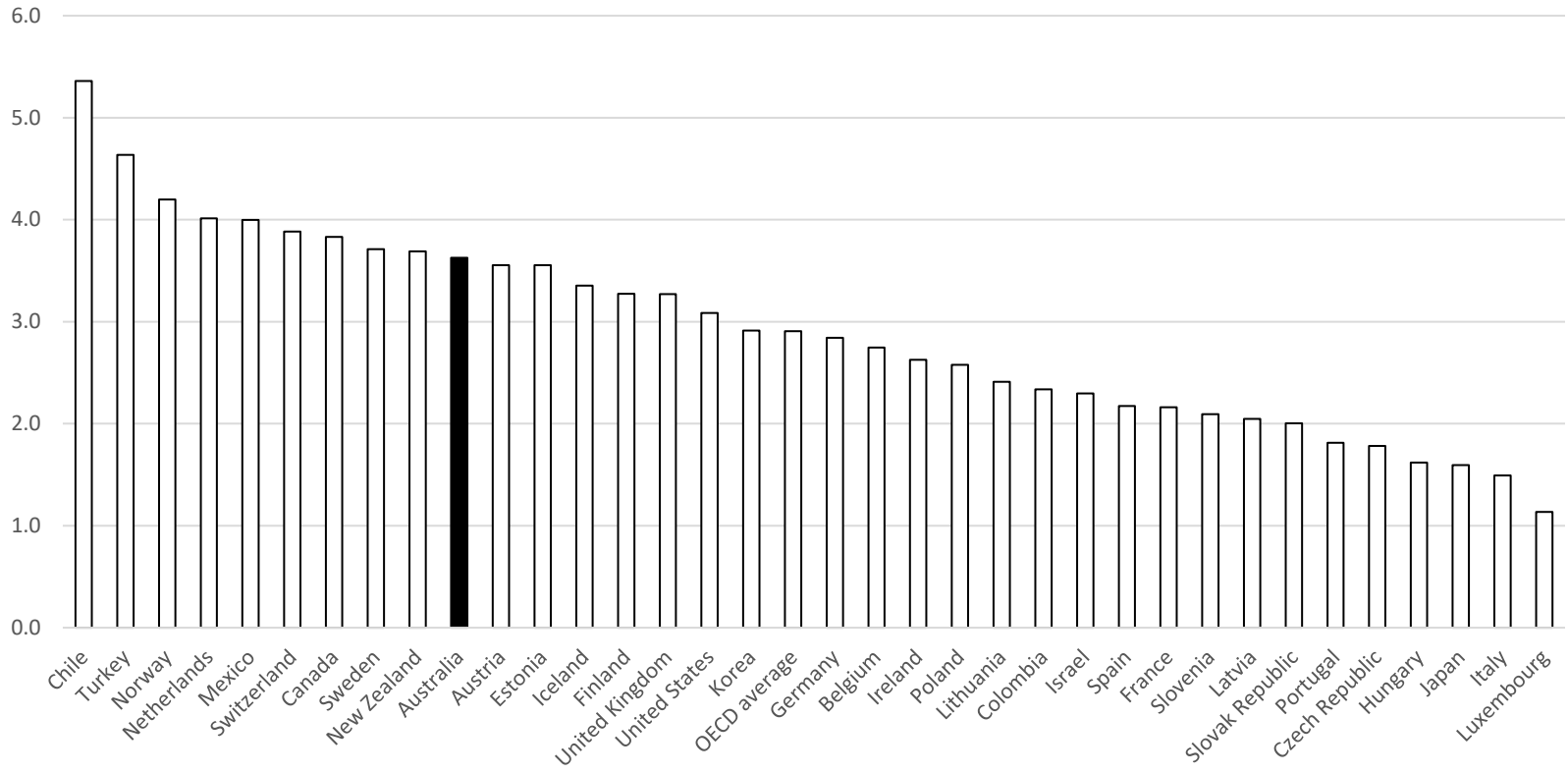
VU began in earliest form as a vocational education and training provider in 1916, in the inner-city suburb of Footscray. After operating for eighty years and growing in student numbers and course diversity, the Dawkins reforms converted VU (and many other institutions) from a College to a University in the 1990s (Croucher et al., 2013; Dawkins, 1987). Per Australia's Excellence for Research in Australia 2018 report (Australian Research Council, 2018), VU research outputs were ranked 'Well above world standard' in the areas of Engineering, Electrical and Electronic Engineering, with notable works in water filtration and informatics, Human Movement and Sports Science, Nursing, Medical Physiology, and Public Health and Health Services; as well as nine other research areas being ranked 'Above world standard'. The Times Higher Education ranks VU in the 301-350 global bracket and the 45th young University (The Times, 2019).

In 2018 VU had an enrolment count of 16,158 domestic and 11,058 international students, ranking 25th in Australia in terms of total enrolments, translating to 1.74% National and 5.73% State market share (Department of Education, 2018a). The student body is culturally and economically diverse, and includes target equity groups that require particular focus in the expansion of the sector. In 2018 VU had the second-highest proportion of low socio-economic students in the State of Victoria (Department of Education, 2018b), with 25.47% classed as low-SES, compared with 7.80% at The University of Melbourne. Other equity indicators, such as

proportion of students from a regional or remote area, do not follow the same institutional profile.

VU employed just under 2000 full time equivalent (FTE) staff in 2018, 27.58% of whom were academic and 58.02% professional. Of the academic staff, 63.04% were tenured/ongoing and the remainder fixed term or casual (VU Annual Report, 2018, p. 33-35). As a proportion of all FTE staff employed in Universities in Victoria in 2018, VU accounts for 4.54% (Department of Education, 2018), which nationally places it 31st out of 43 Australian Universities, in terms of number of staff employed.

Figure SM-C-1. OECD nation proportion of total Government expenditure on tertiary education in 2016.



Appendix D: Sampling details, distributional and imputation characteristics

Table SM-D-1. Sample size across years.

Control variable	2016 (traditional mode) *n* (%) / *M*(*sd*)	2017 (traditional mode) *n* (%) / *M*(*sd*)	2018 (block mode) *n* (%) / *M*(*sd*)
commencing student	22351 (79.73%)	21597 (76.26%)	24690 (82.28%)
continuing student	5684 (20.27%)	6725 (23.74%)	5318 (17.72%)
attending full time	27284 (97.32%)	27358 (96.60%)	29074 (96.89%)
attending part time	931 (3.32%)	964 (3.40%)	934 (3.11%)
VTAC applicant	14699 (52.43%)	14887 (52.56%)	16582 (55.26%)
direct applicant	13516 (48.21%)	13435 (47.44%)	13426 (44.74%)
gender, male	13796 (49.21%)	13300 (46.96%)	13352 (44.49%)
gender, female	14419 (51.43%)	15022 (53.04%)	16656 (55.51%)
primary language, English	19519 (69.62%)	19270 (68.04%)	19693 (65.63%)
primary language, not English	8347 (29.77%)	8785 (31.02%)	10173 (33.90%)
primary language, missing	349 (1.24%)	267 (0.94%)	142 (0.47%)
country of birth, Australia	20774 (74.10%)	20474 (72.29%)	21448 (71.47%)
country of birth, not Australia	7437 (26.53%)	7841 (27.69%)	8558 (28.52%)
country of birth missing	4 (0.01%)	7 (0.02%)	2 (0.01%)
ATSI	177 (0.63%)	231 (0.82%)	220 (0.73%)
Not ATSI	28010 (99.91%)	28056 (99.06%)	29611 (98.68%)
ATSI missing	28 (0.10%)	35 (0.12%)	177 (0.59%)
disability	1707 (6.09%)	1670 (5.90%)	2017 (6.72%)
no reported disability	26508 (94.55%)	26652 (94.10%)	27991 (93.28%)
region, Western suburbs	13402 (47.80%)	13897 (49.07%)	14625 (48.74%)
Region, Other region in Victoria	13924 (49.67%)	13427 (47.41%)	14132 (47.09%)
Region, Overseas	826 (2.95%)	923 (3.26%)	1154 (3.85%)
Region, missing	63 (0.22%)	75 (0.26%)	97 (0.32%)

ATAR	*M* = 53.23(16.51)	M = 53.38(16.86)	M = 51.57(16.67)
ATAR missing	6406 (22.85%)	6428 (22.70%)	6531 (21.76%)
AGE	M = 21.83(6.35)	M = 21.80(6.15)	M = 21.46(5.95)
SEIFA	M = 54.01(28.29)	M = 54.15(27.92)	M = 53.64(27.79)
SEIFA missing	898 (3.20%)	999 (3.53%)	1237 (4.12%)
Discipline: Health and Biomedicine	7489 (26.71%)	9199 (32.48%)	11739 (39.12%)
Discipline: Arts and Education	6607 (23.57%)	7010 (24.75%)	6712 (22.37%)
Discipline: Business	5503 (19.63%)	4105 (14.49%)	3388 (11.29%)
Discipline: Sports and Exercise Science	3993 (14.24%)	3462 (12.22%)	3990 (13.30%)
Discipline - Law and justice	2614 (9.32%)	2513 (8.87%)	1878 (6.26%)
Discipline - Engineering and Science	2009 (7.17%)	2033 (7.18%)	2301 (7.67%)
Postgraduate	4266 (15.22%)	4459 (15.74%)	4702 (15.67%)
Bachelor	6107 (21.78%)	6517 (23.01%)	6664 (22.21%)
VET	4333 (15.46%)	4388 (15.49%)	4383 (14.61%)
Secondary school	5833 (20.81%)	5627 (19.87%)	6447 (21.48%)
Pre-secondary school	4812 (17.16%)	4885 (17.25%)	5486 (18.28%)
missing	2864 (10.22%)	2446 (8.64%)	2326 (7.75%)
Total n	28035	28322	30008

Note. For continuous variables a mean with 95% confidence interval and standard deviation is reported. SEIFA = Socio-Economic Indexes for Areas. Commencing refers to students who have no unit enrolment attempt prior to any in that year of study (i.e. a new student at VU). First in family is derived from the highest education level of either parent.

Table SM-D-2. Response rates to satisfaction surveys by discipline and year.

Discipline	2016	2017	2018	Change in Block
Health and Biomedicine	15.98%	18.10%	18.90%	1.86%
Arts and Education	21.17%	21.76%	20.25%	-1.22%
Business	19.40%	14.06%	16.20%	-0.53%
Sport and Exercise Science	18.54%	17.73%	19.51%	1.38%
Law and Justice	19.48%	16.57%	19.12%	1.09%
Engineering and Science	11.92%	11.65%	23.16%	11.38%

Note. Change in Block = Difference between average response rate of 2016 and 2017, and the response rate in 2018 (introduction of block mode). Flag for completion is response to any SET or SEU item.

Table SM-D-3. Distributional characteristics of the outcome variables.

	M (95% IC)	*SD*	Skewness	Kurtosis
SEU1	3.99 [3.98, 4.01]	1.03	-1.08	0.77
SEU2	4.03 [4.02, 4.05]	1.01	-1.08	0.81
SEU3	4.06 [4.04, 4.07]	0.99	-1.12	1.04
SEU4	4.03 [4.02, 4.05]	1.01	-1.10	0.84
SEU5	4.15 [4.13, 4.16]	0.91	-1.19	1.57
SEU6	4.04 [4.02, 4.05]	1.00	-1.13	1.01
SET1	4.25 [4.24, 4.27]	0.96	-1.49	2.00
SET2	4.11 [4.10, 4.13]	1.00	-1.16	0.91
SET3	4.09 [4.08, 4.11]	1.06	-1.19	0.84
SET4	4.20 [4.18, 4.21]	0.97	-1.28	1.29
SET5	4.06 [4.05, 4.08]	1.03	-1.07	0.64
SET6	4.24 [4.23, 4.26]	0.94	-1.41	1.79

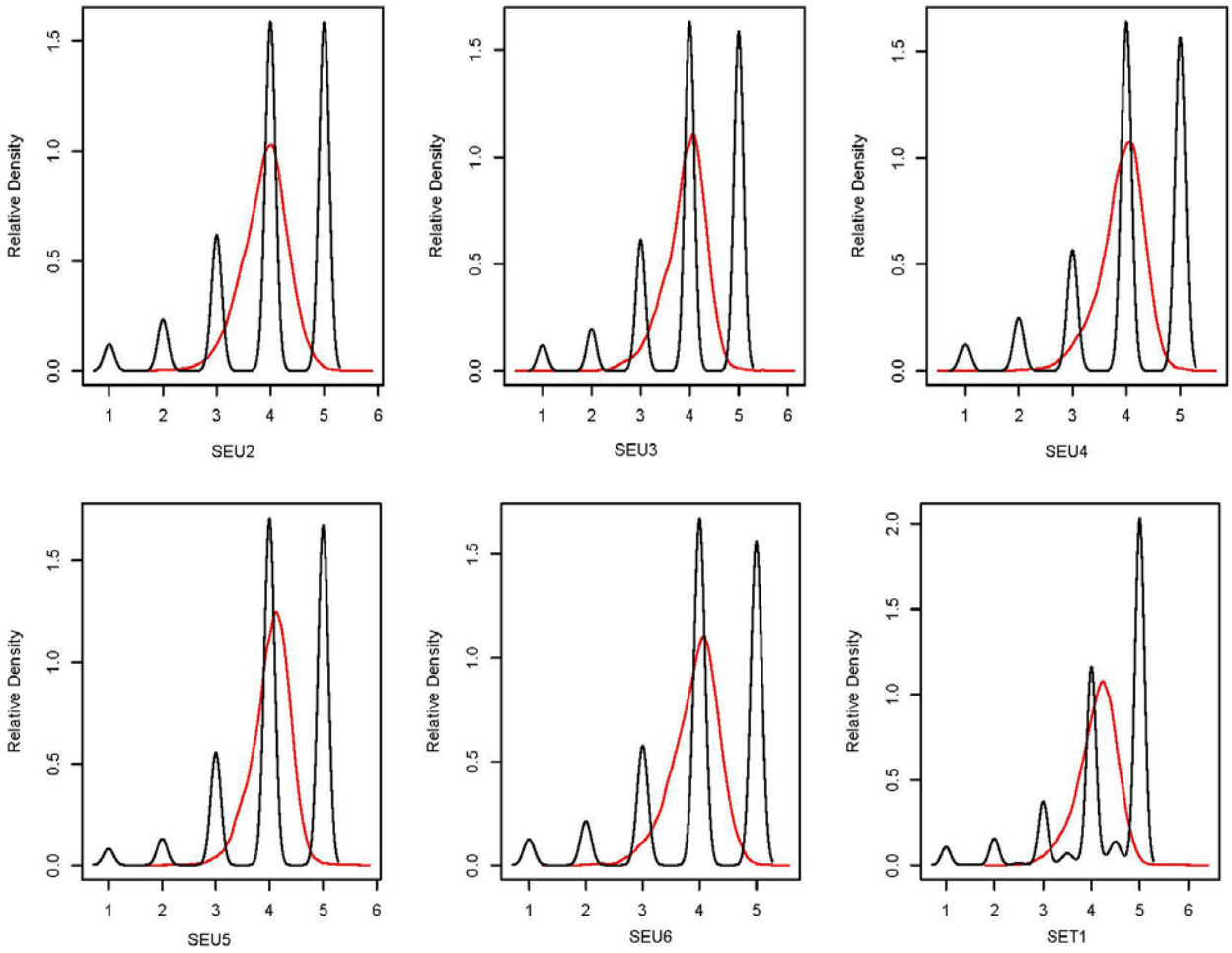
Note. SET and SEU items range in score from 1 (strongly disagree) to 5 (strongly agree).

Table SM-D-4. Outcome variable mean comparison across traditional and block mode.

Outcome	Traditional mode	Block mode	Mean difference	Welsch Independent samples t-test
	M [95 CI] (*SD)*	*M* [95 CI], (*SD)*	*M [95 CI]*	*T(df)*
SEU1	4.00 [3.98, 4.02] (1.01)	3.98 [3.95, 4.01] (1.06)	-0.02 [-0.01, 0.05]	-1.70(11631)
SEU2	4.03 [4.01, 4.05] (0.98)	4.04 [4.02, 4.07] (1.01)	0.01 [-0.04, 0.02]	0.69(11744)
SEU3	4.06 [4.04, 4.07] (0.98)	4.06 [4.03, 4.08] (1.00)	0.00 [-0.03, 0.02]	-0.06(11743)
SEU4	4.03 [4.01, 4.05] (0.99)	4.03 [4.00, 4.05] (1.02)	-0.00 [-0.03, 0.04]	0.44(11656)
SEU5	4.14 [4.12, 4.15] (0.90)	4.17 [4.15, 4.20] (0.91)	0.03 [-0.07, 0.01]	2.41(11698) *
SEU6	4.09 [4.07, 4.11] (0.93)	3.95 [3.92, 3.98] (1.10)	-0.14 [0.10, 0.17]	-7.99(10286)***
SET1	4.21 [4.19, 4.23] (0.97)	4.34 [4.31, 4.36] (0.92)	0.15 [0.18, 0.12]	9.59(12402)***
SET2	4.05 [4.03, 4.07] (1.01)	4.23 [4.20, 4.25] (0.94)	0.20 [0.23, 0.16]	12.12(12682)***
SET3	4.04 [4.02, 4.06] (1.06)	4.19 [4.16, 4.22] (1.01)	0.17 [0.21, 0.14]	10.04(12444)***
SET4	4.12 [4.10, 4.14] (0.99)	4.33 [4.30, 4.35] (0.90)	0.09 [0.12, 0.06]	5.25(11675)***
SET5	4.00 [3.98, 4.02] (1.04)	4.17 [4.15, 4.20] (0.98)	0.19 [0.22, 0.16]	11.43(12523)
SET6	4.21 [4.19, 4.23] (0.94)	4.33 [4.29, 4.34] (0.91)	0.13 [0.16, 0.10]	8.21(12417)***
Marks	56.51 [56.30, 56.72] (24.91)	63.54 [63.27, 63.81] (23.37)	7.03 [7.37, 6.69]	40.13(62089)***

Note. For these inferential tests years 2016 and 2017 are combined and the mean compared to 2018.

Figure SM-D-1. Distributions of original and imputed data.



Note. Red distributions denote imputed values.

Appendix E: Intra-class correlations demonstrating nested structures in the data.

Table SM-E-1. ICCs and random intercept model comparison for all outcome variables

	Random intercept for unit				Random intercept for student				Random intercept for teacher			
Item	Random intercept SD	Random residual	ICC	Deviance $\chi2(1)$	Random intercept SD	Random residual	ICC	Deviance $\chi2(1)$	Random intercept SD	Random residual	ICC	Deviance $\chi2(1)$
SEU1	0.31	0.99	0.09	972.31	0.45	0.93	0.19	564.06	0.40	0.96	0.15	1460.13
SEU2	0.29	0.97	0.08	757.57	0.44	0.91	0.19	595.36	0.35	0.95	0.12	1099.30
SEU3	0.27	0.95	0.08	849.94	0.45	0.88	0.21	672.04	0.34	0.93	0.12	1203.96
SEU4	0.28	0.97	0.08	864.46	0.45	0.91	0.2	619.37	0.37	0.94	0.13	1372.47
SEU5	0.24	0.88	0.07	608.27	0.43	0.8	0.23	799.21	0.31	0.86	0.11	928.555
SEU6	0.29	0.96	0.08	881.84	0.48	0.88	0.23	816.55	0.33	0.95	0.11	960.084
SET1	0.27	0.92	0.08	1011.63	0.37	0.89	0.15	371.77	0.46	0.86	0.22	2393.71
SET2	0.26	0.97	0.07	896.21	0.41	0.92	0.17	466.77	0.42	0.92	0.17	1883.78
SET3	0.26	0.97	0.07	896.21	0.43	0.97	0.16	466.58	0.51	0.96	0.22	2294.67
SET4	0.24	0.94	0.06	745.86	0.41	0.88	0.17	526.18	0.39	0.90	0.16	1710.7
SET5	0.27	1.00	0.07	820.34	0.46	0.93	0.2	639.13	0.44	0.95	0.18	1862.24
SET6	0.26	0.91	0.07	875.80	0.37	0.86	0.16	416.06	0.46	0.85	0.23	2275.43
Marks	7.08	23.4	0.08	7918.51	20.42	15.03	0.65	44844.57	6.75	13.93	0.19	2257.35

Note. ICCs represent the proportion of total variance explained by group membership and were calculated per Bliese, 2006, p. 54. Deviance reflects the difference in log likelihood[2] between the multilevel model that estimates variance around the mean, to a mean-only model that does not include a random intercept. All $\chi2$ deviance parameters are significant above $p<.001$. *n* observations = 16,034 for SEU items and *n* = 15,989 for SET items, *n* = 82.031 for marks. Nested level include *n* distinct students = 6325, *n* distinct units = 136 and *n* distinct teachers = 613. Note models were subset to responders of the SET survey in order to calculate ICC parameters for models with a random intercept for teacher (*n* observations = 15,989).

References

Australian Research Council (2018). *State of Australian University Research 2018-19*. https://dataportal.arc.gov.au/ERA/NationalReport/2018/

Baik, C., Naylor, R., & Arkoudis, S. (2015). *The first year experience in Australian Universities: Findings from two decades, 1994-2014*. Melbourne Centre for the Study of Higher Education. https://melbourne-cshe.unimelb.edu.au/__data/assets/pdf_file/0016/1513123/FYE-2014-FULL-report-FINAL-web.pdf

Baik, C., Naylor, R., Arkoudis, S., & Dabrowski, A. (2019). Examining the experiences of first-year students with low tertiary admission scores in Australian universities. *Studies in Higher Education*, *44*(3), 526-538. http://doi.org/10.1080/03075079.2017.1383376

Bradley, D., Noonan, P., Nugent, H., & Scales, B. (2008). Department of Education and Training. *Review of Australian higher education: Final report*. http://hdl.voced.edu.au/10707/44384

Croucher, G., Marginson, S., Norton, A., & Wells, J. (2013). *The Dawkins revolution : 25 years on*. Melbourne University Publishing

Coaldrake, P. (2019). *What's in a name? Review of the higher education provider category Standards*. Higher Education Standards Panel. https://docs-edu.govcms.gov.au/node/53071

Dawkins, J. S. (1987). *Higher education: A policy discussion paper*. Australian Government Publishing Service. https://www.voced.edu.au/content/ngv%3A1033

Department of Education and Training (2019a). Selected Higher Education Statistics – 2018 Student data. https://www.education.gov.au/node/26036

Department of Education and Training (2019b). Selected Higher Education Statistics – 2018 Section 11 Equity groups. https://docs.education.gov.au/node/53025

Department of Education and Training (2019c). Selected Higher Education Statistics – Staff Data 2018. https://www.education.gov.au/staff-data

Dobson, I. R. (2012). PhDs in Australia, from the beginning. *The Australian Universities' Review, 54*(1), 94

Dow, C. (2013). *Higher education: Sustainability of a demand-driven system.* Parliament Library Briefing Book – 44th Parliament. https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/BriefingBook44p/HigherEducation

Gascoigne, J. (1996). The cultural origins of Australian universities. *Journal of Australian Studies, 20*(50-51), 18-27

Harvey, A., Szalkowicz, G., & Luckman, M. (2017). *The re-recruitment of students who have withdrawn from Australian higher education.* Centre for Higher Education Equity and Diversity Research. La Trobe University Press

Higher Education Standards Panel. (2017). *Final report: Improving retention, completion and success in higher education.* https://docs.education.gov.au/node/50816

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

Naylor, R., Baik, C. & Arkoudis, S. (2018). Identifying attrition risk based on the first year experience. *Higher Education Research & Development, 37*(2), 328-342. doi: 10.1080/07294360.2017.1370438

Organisation for Economic Co-operation and Development (OECD). Education at a Glance 2019 – Data and Methodology (Indicator C3). Accessed 21/11/2019, from https://www.oecd.org/education/education-at-a-glance/educationataglance2019-dataandmethodology.htm

Quality Indicators of Learning and Teaching (2017). Student Experience Survey, 2017 National Report. Available at: https://www.qilt.edu.au/docs/default-source/ses/ses2018/2018-ses-national-report75e58791b1e86477b58fff00006709da.pdf?sfvrsn=d733e33c_4

Tertiary Education Quality and Standards Agency (2019). National Register. https://www.teqsa.gov.au/national-register

The Times Higher Education (2019). World University Rankings. https://www.timeshighereducation.com/world-university-rankings/2020/

Victoria University (2018). Annual Report. https://www.vu.edu.au/sites/default/files/vu-annual-report-2018.pdf