Students' Conceptions of Bell Curve Grading Fairness in Relation to Goal Orientation and Motivation

Lynette TAN Yuen Ling
National University of Singapore, rc4lynette@nus.edu.sg

Brenda Yuen
National University of Singapore, elcbypl@nus.edu.sg

Wee Ling LOO
Singapore Management University, wlloo@smu.edu.sg

Christiaan Prinsloo
Seoul National University, christiaan@snu.ac.kr

Mark Gan
National University of Singapore, mark.gan@nus.edu.sg

Recommended Citation
TAN Yuen Ling, Lynette; Yuen, Brenda; LOO, Wee Ling; Prinsloo, Christiaan; and Gan, Mark (2020)
"Students' Conceptions of Bell Curve Grading Fairness in Relation to Goal Orientation and Motivation,"
International Journal for the Scholarship of Teaching and Learning: Vol. 14: No. 1, Article 7.
Available at: https://doi.org/10.20429/ijsotl.2020.140107

Students' Conceptions of Bell Curve Grading Fairness in Relation to Goal Orientation and Motivation

Abstract

The controversial bell curve has received considerable attention in recent years as a grade distribution tool where "norm-referenced grading involves comparing students' performances with each other" rather than where they fall on a "predefined continuum of quality" (Brookhart, 2013, p. 258). Despite educators' deep concern on the fairness of bell curve grading, there is little research done on students' conceptions of that grading system in higher education. This correlational study uses open-ended questions and three instruments to measure students' conceptions of the fairness of bell curve grading, their goal orientations, and motivation. Undergraduates from three universities participated in the survey (N= 211). Results suggest that students have a formalized conception of bell curve grading, perceive it to be generally fair, but tend to hold negative views about its impact on learning. The correlations with their goal orientation and levels of motivation, while yielding constructive inferences, were not overly significant.

Keywords

Bell curve grading, fairness, goal orientation, motivation, tertiary education

Creative Commons License

Cover Page Footnote

Students' Conceptions of Bell Curve Grading Fairness in Relation to Goal Orientation and Motivation

Lynette TAN Yuen Ling[1], Brenda YUEN Pui Lam[1], Wee Ling LOO[2], Christiaan Prinsloo[3], and Mark Gan[1]

1 National University of Singapore
2 Singapore Management University
3 Seoul National University

Abstract

The controversial bell curve has received considerable attention in recent years as a grade distribution tool where "norm-referenced grading involves comparing students' performances with each other" rather than where they fall on a "predefined continuum of quality" (Brookhart, 2013, p. 258). Despite educators' deep concern on the fairness of bell curve grading, there is little research done on students' conceptions of that grading system in higher education. This correlational study uses open-ended questions and three instruments to measure students' conceptions of the fairness of bell curve grading, their goal orientations, and motivation. Undergraduates from three universities participated in the survey (N= 211). Results suggest that students have a formalized conception of bell curve grading, perceive it to be generally fair, but tend to hold negative views about its impact on learning. The correlations with their goal orientation and levels of motivation, while yielding constructive inferences, were not overly significant.

INTRODUCTION

The *Encyclopedia of Educational Theory and Philosophy* (2014) credits the 16th century for the invention of the bell curve. The bell curve (i.e., Gaussian curve or normal distribution) suggests that the statistical distribution of elements is a natural phenomenon that is highly probable and therefore normative. In education, this means that most students obtain average or "normal" grades and relatively few excel and/or fail (Fendler & Muzaffar, 2008). Being statistically assigned into a grading curve spurred students in Singapore to entrust an elusive "Bell Curve God", who resides in cyberspace, with the fate of their grades lying beyond their own efforts or hard work (http://nus-bell-curve-god.appspot.com/mainpage). The president of the National University of Singapore (NUS), Professor Tan, explained bell curve grading "as a tool to moderate grades, and as a guide to prevent grade inflation or deflation." (Tan, 2012). Singapore Management University (SMU) then provost Professor Kong revealed that the "Bell Curve God" is a coping strategy with which students desperately attempt to mitigate the anxiety induced by grading along the curve (Kong, 2016). The sense of student desperation and helplessness to entrust a cyber god with grading fairness together with a literature review on the contentiousness of the bell curve in education provide the primary motivation for the present study. Our purpose is to investigate whether students' conceptions of the fairness of bell curve grading correlate with their goal orientation and motivation. To this end, we pose two research questions:

1. What are students' conceptions of the fairness of bell curve grading in terms of interactional fairness, procedural fairness and fairness of outcome?

2. To what extent do students' conceptions of bell curve grading fairness relate to their goal orientation and motivation?

LITERATURE REVIEW

In the present paper, *bell curve grading* refers to a predetermined curve (not necessarily bell-shaped, possibly adjusted) in which students' grades are impacted by their placement in an overall predetermined distribution pattern where raw scores are converted to acquire a desired mean. This is also referred to as *grading on a curve*, *norm-referenced* or *relative grading* that is governed by particularistic rules. The opposite side of the grading spectrum (governed by meritocratic rules) includes self-referenced, criterion-referenced, and/or absolute grading where students obtain grades based on a "predefined continuum of quality" (Brookhart, 2013, p. 258).

The bell curve has been applied to various natural and social scientific and humanities-related contexts. Perhaps one of the most controversial modern social scientific applications is Herrnstein and Murray's (1994) *The bell curve: Intelligence and class structure in American life*. Part of the controversy is based on the correlation that the authors drew between race and intellect on a statistical curve that designated certain races as inept for education. Without engaging the political ramifications of Herrnstein and Murray's book, the same statistical principle has been applied to higher education with many scholars calling for its abolition in both (Bersin, 2014; Ganguly, 2015; Kong, 2016; Mcgregor, 2013; Nelson, 2011; Tan, 2012; Yount, 2011).

The bell curve portrays a dominant educational discourse that assumes success or failure in learning can be normalized - a sacred normal-as-average viewpoint. Grading on a normal curve is often justified in terms of the application of an absolute statistical truth (i.e., an unchangeable, formalized and normalized concept) of a large population to the ever-changing vicissitudes of education (Fendler & Muzaffar, 2008). Teachers and students who subscribe to this view, either consciously or unconsciously, might fall into the trap of a fixed mindset (Klapp, 2015) where they believe that one is born with a fixed amount of intelligence and ability as

determined by a normal distribution and adopts failure-avoiding behaviors in learning. Bloom (1971) argues that bell-curve thinking and the consequential emotional impact of failure are detrimental to students' feelings and motivation to learn.

Despite its negative impacts, bell curve grading remains a tenacious practice at universities around the world where the "normal curve is seen as the 'silent partner' of the grading system" (Brookhart et al, 2016, p. 832). Indeed, its tenacity may be ascribed to the notion that normal distribution is "generally regarded to be a fact of life" (Fendler & Muzaffar, 2008, p. 64). In higher education, grading on a curve is typically justified as a means to curb grade inflation and deflation, to distinguish highly competitive student populations, to maintain normal distribution over large student cohorts and to preserve institutional reputations (see Close, 2009; Czibor, Onderstal, Sloof, & Van Praag, 2014; Grant, 2016; Kulick & Wright, 2008; Tan, 2012).

Against the background of Herrnstein and Murray's controversy, grade manipulation at universities has elicited much pedagogic scholarship that theorizes grading on a curve in relation with fairness. For example, perceptions of fairness mediate the relationship between grades and student-teacher evaluations (Wendorf & Alexander, 2005). Fairness increases students' satisfaction with education and improves achievement while lack of fairness has contributed to "poor achievement, to attrition, and even to campus vandalism" (Rodabaugh, 1996, p. 37). Scholarship distinguishes the following three highly correlated kinds of fairness or dimensions of teacher credibility (Chory, 2007) that illuminate students' conceptions of grading on a curve, *viz.* interactional, procedural and outcome fairness. These dimensions of fairness were adopted in our data collection protocol to examine the relationships between bell curve grading fairness, and goal orientation and motivation.

Interactional fairness is concerned with interpersonal exchanges (Wendorf & Alexander, 2005). An interactional injustice occurs when students perceive teachers' favouritism toward other students based on gender, race or age and/or teachers who demonstrate "angry" or "mean" behaviour (Rodabaugh, 1996, p. 38). Scholars who defend grading on a curve maintain that it differentiates student achievement, and thus distinguishes students who may enter practice from those who continue to graduate school (Kulick & Wright, 2008). However, Czibor et al (2014) consistently found that men respond more positively to grading on a curve than women, and this gender variance is perceived as an interactional injustice by students when grading on a curve is employed by the teacher.

Procedural fairness is concerned with the methods involved in obtaining and calculating grades, such as tests that accurately measure learning (Rodabaugh, 1996; Wendorf & Alexander, 2005). Proponents of grading on a curve argue that it manages idiosyncratic grading practices where teachers are either lenient or strict, or examinations are too easy or difficult (O'Halloran & Gordon 2014; Tan, 2012; Redding 1998; Weil & Kroontje, 1977). The conception of procedural fairness is supported by the assumption that properly designed, norm-referenced, multiple choice tests render normally distributed results (Kulick & Wright, 2008). However, in classes where student performance is relatively similar, such as within the highly competitive universities included in the sample, luck seems to support normal distribution, and this may discourage students.

Fairness of outcome (i.e., distributive fairness) indicates how accurately grades reflect students' contributions to class and performance on assignments (Rodabaugh, 1996; Wendorf & Alexander, 2005). Curved grading infringes fairness of outcome through the rationale that it prevents grade inflation or deflation (Tan, 2012). For example, if the curve allows for 20% As, 40% Bs, and 40% Cs, then the fairness of outcome is infringed for a student who deserves an A yet receives a B because of the curve. The predetermined manipulation of grades assumes that high academic achievement is a scarce resource for which students compete hence creating an "atmosphere that's toxic by pitting students against one another" (Grant, 2016, n.p.). This counters the benefits of collaborative learning. In terms of distributive justice (i.e., the fair distribution of scarce resources), grading on a curve seems illogical because grades (A through F) are not scarce resources. Designing courses and grading systems on the premise that certain grades were "scarce goods" would be ethically questionable (Close, 2009, p. 365). Further, if educational efforts were successful, the achievement distribution should differ from a normal curve. Forcing a normal distribution would then infringe fairness of outcome (Bloom, 1971).

The use of bell curve grading at the National University of Singapore (NUS), Singapore Management University (SMU) and Seoul National University (SNU)

This section describes the different ways and the extent to which norm-referencing is used in particular modules taught at the three universities from which data were collected. Beginning with NUS a bell curve is used in the assessment of level 1 and 2 *Ideas and Exposition* modules and *Basic English* modules. The *Ideas and Exposition* module involves students thinking critically, understanding the rhetorical principles of composition, and eventually writing persuasively. The grade distribution tool is employed in two categories: first, at the individual module (ranging between 52-60 students), and again at the entire level (350 students). The use of the bell curve in the programme follows what was described by Tan (2012). Professor Tan asserts that the bell curve is flexibly employed as follows: (i) it is employed only where the class size is large enough (preferably above 30); (ii) small deviations from the norm-based grade distribution guidelines are ignored; (iii) instructors with strong justification may deviate from the guidelines; and (iv) high cumulative grade point average profiles of the class could justify such deviation. In this module, the class size (above 30) fits the first caveat. Its use is also appropriate to maintain fairness, since the instructors teaching the modules hail from all over the world, mainly America, South Africa, Britain, Australia, India and Singapore, and grading styles and leniency or strictness in grading could differ. In the modules, extensive scaffolding, support and feedback are provided to the students to refine their final papers. Thus, as with an honours class with a high cumulative grade point average profile, the curve is adjusted upwards. The results are then viewed against results across faculties and schools to monitor occurrences of grade inflation or deflation.

The *Basic English* module is a required academic writing module that is offered to undergraduate students who do not achieve the required standard of the university's English writing placement test. The module aims to enhance students' English language skills in reading, writing and grammar in order to help them meet their language needs for university. To facilitate student

learning, the class size is limited to a maximum of 15 students with a cohort size varying from 60 to 130 students per semester. This module implements an outcomes-based approach to teaching and learning (Biggs & Tang, 2011) which requires clearly stated assessment criteria aligned to the declared outcome, i.e., it requires criterion-referenced assessment of student performance. As NUS imposes university-wide normative grade distribution guidelines, the module adopts a hybrid model to assessment, to accommodate both norm and criterion referencing in assessment. The hybrid model involves three elements, norm referencing, criterion referencing and actual grading, in a feedback loop which can be iterated many times for moderation to achieve expected distribution of grades after a few cycles of discrepancies (Lok, McNaught & Young, 2016).

At SMU, the *Ethics and Social Responsibility* module is a course in applied ethics. It aims to create awareness and sensitivity to ethical issues in diverse contexts, and to equip students with the ability to describe and analyse the issues, to come to reasoned and persuasive conclusions. The module is a graduation requirement, and class size is usually limited to a maximum of 45 students. When data was collected, there were 8 classes, divided among 5 instructors, of 41 to 45 students per class, bringing the cohort to a total of 350. SMU takes account of absolute grades as an indication of performance, but has limited norm-based grade distribution guidelines in place to ensure no grade inflation or deflation, or overly lenient or strict marking across instructors. Regarding the *Ethics* module, there are recommended guidelines for the A-grade category alone; these are employed flexibly similar to NUS. The guidelines were applied across 128 students taught by the same instructor. The instructors of the module are from different countries (e.g., North America, Taiwan, Sri Lanka and Singapore) further justifying the need to ensure some parity in grading. The module coordinator reviews each instructor's grade distribution and adjusts for unusual variance among the different instructors. All grade distributions then undergo another round of checking at the school level and finally across all schools.

At SNU, the *College English Program* comprises four modules of tiered – foundational, intermediate (lower and upper), and advanced – English courses into which undergraduate students are divided based on their English language proficiency and individual departmental/faculty requirements. As a foreign language graduation requirement, all undergraduate students must complete at least one module, with the exception of exempted students. Data was collected from the lower intermediate and advanced modules entitled *College English* and *Advanced English*.

College English modules are limited to a maximum of 20 students who are graded on a strict bell curve that requires the following grade distribution: a maximum of 30% As, 40% Bs, and 30% Cs or lower. *Advanced English* modules included in this study were capped at between 13-17 students. These modules are graded on an inverted V-curve that allows for a maximum of 50% A-grades and 50% B-grades or lower. These relatively rigid grading norms were implemented to curb grade inflation. Different from the other universities in the sample, classes within modules in the *College English Program* at SNU are not integrated to apply the bell curve over larger cohorts. The curve is strictly applied per class, and class size is generally limited to a maximum of 20 students. This means that the bell curve is currently applied to cohorts that are approximately 1/3 smaller than what Tan (2012) advises. While *College English* grades are curved strictly, the bottom 30% of students is given the opportunity to retake the module in a follow-up semester with another instructor. At the time of writing this paper, the *College English Program* was deliberating the implementation of absolute grading across the entire program.

METHOD
Participants
A cohort of 211 undergraduates from three universities participated in the study. NUS participants were divided into two groups: those taking a compulsory English language proficiency module *Basic English* (N=41), and an elective *Ideas and Exposition* module, *Women in Film* (N=58) where students are part of a residential programme and major in a variety of academic disciplines. Bell curve grading was used in the former group, while an adjusted bell curve was implemented in the latter. SNU participants consisted of undergraduates taking *College English* (N=24) and *Advanced English* (N=10). Both groups took English as a compulsory graduation requirement. *College English* grades were bell curved and *Advanced English* grades were distributed on an inverted V-curve. Both modules were attended by students from various disciplines across the humanities and social and natural sciences, rendering an interdisciplinary cohort. SMU participants took a compulsory module *Ethics and Social Responsibility* (N=78). They were from 3 classes taking the module under the same instructor. Guidelines pertaining to the A grade category alone applied. Similar to SNU, students from various disciplines attended the classes (see Table 1).

Table 1. Types of modules					
	Language modules				Content-based module
University	NUS	NUS	SNU	SNU	SMU
Module title	Basic English	Women in Film	College English	Advanced English	Ethics & Social Responsibility
Module type	Academic English Writing	Expository Writing	College English	Advanced English (Academic English/ World Literature)	Ethics & Social Responsibility
Prerequisite	Below the required standard of the university-wide English writing placement test	TOEFL iBT (114 and above) / Band 3 in the university-wide English writing placement test	TOEFL iBT (55-65)	TOEFL iBT (80 and above)	No
Class size	15	15	18-19	13-17	40-45
Cohort size	66	69	37	30	128
Participants	41	58	24	10	78
Credit bearing module	No	Yes	Yes	Yes	Yes
Compulsory/ elective module	Compulsory	Elective	Compulsory	Compulsory	Compulsory
Bell curve grading	Bell curve grading	Adjusted bell curve grading	Bell curve grading	Adjusted bell curve grading	Adjusted bell curve grading

MEASURES

The instrument, which comprises three main sections, has 53 items weighted on a 7-point Likert scale ranging from *not true of me* (1) to *extremely true of me* (7), and two open-ended questions that require students to describe their personal experiences of any negative or positive effects from the use of bell curve grading, and whether they think that the use of bell curve grading improves or worsens their university's grading practices.

The first section ascertains students' conception of the fairness of the bell curve as a grading strategy. The items used in this section were adapted from Rodabaugh (1996) and Gordon and Fay's (2010) measures of fairness in college teaching. Nine statements were written to assess the three aspects of the fairness of bell curve grading, namely *interactional fairness, procedural fairness* and *fairness of outcomes. Interactional fairness* was measured by statements written about the provision of equal access to grading information when bell curve grading is used ("When bell curve grading is used, teachers provide clear explanation of how they grade my assignments, tests or exams."); *procedural fairness* was assessed with statements about the accuracy and consistency of bell curve grading ("I think that my teachers within and across departments are more consistent in grading when they apply the bell curve."). To measure the *fairness of outcomes*, participants were asked whether grades match their learning when bell curve grading is applied (e.g. "When bell curve grading is use, I think that I receive grades that are reflective of what I have learnt."). Students' responses were averaged to compute the indices of the overall fairness and the three aspects of fairness.

The second section comprises Elliot, Murayama and Pekrun's (2011) '3 x 2 achievement goal model'. This model encompasses 6 goal constructs of *task approach* ("My focus is to know the right answers to the questions in assessments."), *task-avoidance* ("My goal is to avoid getting a lot of questions wrong in assessments."), *self-approach* ("My goal is to do better in assessments than I typically do in this type of situation."), *self-avoidance* ("I aim to avoid performing poorly in assessments compared to my typical level of performance."), *other-approach* ("I aim to do better than my classmates in assessments."), and *other-avoidance* ("My aim is to avoid doing worse than other students in assessments."). Students were shown 18 statements that represent types of goals that they use and were instructed to indicate how true each statement was of them. Their responses were averaged to compute the indices for the achievement goal and 6 specific types.

The last section, which is based on Deci, Eghrari, Patrick and Leone's (1994) theory of self-determination, measures students' levels of motivation. 23 items were used to assess intrinsic motivation of the participants for the module they attended. The four constructs of intrinsic motivation include: *interest/enjoyment* ("I enjoyed doing this module very much."); *perceived competence* ("After working at this module for a while, I felt pretty competent."); *effort/importance* ("I put a lot of effort into this module."); and *pressure/tension* ("I was very relaxed in doing this module.").

Scores were averaged to compute the indices for the intrinsic motivation and each specific construct.

Procedure

The survey was conducted during the end of semester 1 of AY2017/2018 at SNU and the end of semester 2 of the same academic year at NUS and SMU. The survey link was released to students via an online platform with participation being anonymous and voluntary. Students were given 3 weeks to complete the survey.

RESULTS
Fairness, Goal Orientation and Motivation scales

Table 2 presents the descriptive statistics and internal consistencies of *fairness, achievement goal* and *motivation* scales. Cronbach's alpha for the *fairness, achievement goal* and *motivation* scales were 0.68, 0.95 and 0.84 respectively. The mean score for the *fairness* scale has the lowest of 3.43 on a 7-point response continuum, implying that students perceived the bell curve grading as a slightly/moderately fair instrument for grading their performances.

Table 2. Descriptive statistics and internal consistencies for the Fairness, Goal Orientation and Motivation scales (N=211)

	Items	Mean	SD	Observed range	Cronbach's alpha
Fairness	9	3.43	0.23	3.05-3.83	.68
Goal Orientation	18	5.36	0.63	4.94-5.68	.95
Motivation	23	4.53	0.69	3.07-5.67	.84

Sub-dimensions of Fairness scale

Table 3 presents the descriptive statistics of the three constructs for the *fairness* scale. Tests for reliability suggested modest reliabilities for the *interactional fairness, procedural fairness* and *fairness of outcomes* scales, with Cronbach's alpha ranging from 0.61 to 0.69. The modest reliabilities reflect the fact that students' responses to these scales may stem from their varied backgrounds and learning experiences given that dissimilar modules were taught by different instructors. They may also be attributable to differing ideas that students have about what a bell curve is and the form of norm referencing that is used in their academic institutions. The overall mean score for *procedural fairness* was the highest at 3.62 while the lowest was *fairness of outcomes* with a score of 3.30.

In terms of how the three institutions fared, the *fairness* score for SNU was the highest with a score of 3.53 (*interactional fairness* was the highest faring construct at 3.93) while SMU had the lowest at 3.35 (*interactional fairness* was the lowest at 2.97). The highest mean score in terms of the three constructs was for SNU at 3.93 for *interactional fairness*.

The three institutions were at a variance in terms of which aspect of grading fairness was deemed to be most affected by bell curve grading. Fairness of outcomes (3.23) acquired the lowest mean score quantitatively for NUS, procedural fairness (3.27) for

Table 3. Descriptive statistics of the sub-dimensions of Fairness scale by institutions

	Overall (N=211)		NUS (n=99)		SNU (n=34)		SMU (n=78)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Fairness	3.43	0.23	3.46	0.87	3.53	0.83	3.35	0.88
Interactional fairness	3.38	0.15	3.33	1.37	3.93	1.43	2.97	1.33
Procedural fairness	3.62	0.26	3.82	1.21	3.27	1.12	3.51	1.25
Fairness of outcomes	3.30	0.71	3.23	1.21	3.29	1.08	3.32	1.30

Table 4. Descriptive statistics and internal consistencies for the sub-dimensions of Goal Orientation scale (N=211)

Goal Orientation	Explanation	Items	Mean	SD	Observed range	Cronbach's alpha
Task-approach	"Do the task correctly."	3	5.58	0.51	5.40-5.68	.87
Self-approach	"Do better than before."	3	5.49	0.35	5.42-5.61	.88
Other-approach	"Do better than others."	3	5.12	0.51	4.94-5.22	.91
Task-avoidance	"Avoid doing the task incorrectly."	3	5.34	0.58	5.18-5.54	.78
Self-avoidance	"Avoid doing worse than before."	3	5.43	0.00	5.42-5.45	.75
Other-avoidance	"Avoid doing worse than others."	3	5.17	0.03	5.16-5.20	.93

SNU and interactional fairness (2.97) acquired the lowest mean score for SMU.

Sub-dimensions of Goal Orientation scale
As shown in Table 4, tests for reliability suggested a fairly high level of reliability for the six types of *goal orientation* with Cronbach's alpha ranged from 0.75 to 0.92. In terms of the six constructs, the highest mean was the *task-approach* goal orientation with a score of 5.58, and the lowest was the *other-approach* goal orientation with a score of 5.17.

Sub-dimensions of Motivation scale
Table 5 displays the descriptive statistics and internal consistencies for the four constructs of *intrinsic motivation*. Tests for reliability suggested a fairly high level of reliability for the four constructs with Cronbach's alpha ranged from 0.80 to 0.89. In terms of the 4 constructs, the highest mean was the *interest/enjoyment* with a score of 5.58, and the lowest was the *perceived competence* with a score of 3.91.

Table 5. Descriptive statistics and internal consistencies for the sub-dimensions of Motivation scale (N=211)

	Items	Mean	SD	Observed range	Cronbach's alpha
Interest/Enjoyment	7	5.23	0.38	4.45-5.67	.89
Perceived competence	5	3.91	0.38	3.55-4.52	.79
Effort/Importance	4*	4.85	0.27	4.48-5.07	.85
Pressure/Tension	5	4.26	0.32	3.97-4.73	.80

Note: *Cronbach's alpha was .44 before item Q11_C_7 was deleted

Correlations of fairness with goal orientation and motivation
Spearman's rho was used to study the linear relationship between students' conception of fairness of the bell curve grading and achievement goal orientations since a significant Shapiro-Wilk statistic (W=0.97, p=0.00) indicates that the data were not normally distributed. As shown in Table 6, the correlations of students' conception of fairness of the bell curve grading with goal orientation by modules were statistically not significant. In terms of specific achievement goals, results indicated positive and moderate correlations between *fairness* and *self-approach* (r_s= 0.41, p<0.05) and *self-avoidance* (r_s= 0.51, p<0.05) goal orientations for SNU's *College English* course*, and other-approach* goal orientation for SNU's *Advanced English* course (r_s= 0.65, p<0.05).

Pearson's correlation coefficient was calculated to study the linear relationship between students' conception of fairness of the bell curve grading and their intrinsic motivation as a non-significant Shapiro-Wilk statistic (W=0.99, p=0.22) indicates that the data were normally distributed. The correlations of students' conception of fairness of the bell curve grading with their level of intrinsic motivation by modules, as shown in Table 6, were statistically not significant. Results indicated a small but negative correlation between *fairness* and *pressure/tension* (r=-0.23, p<0.05) for SMU students. This suggests that the fairer the SMU students perceived bell curve grading was, the less relaxed they were when doing their module.

The qualitative data was acquired from the responses to two open-ended questions that were included in the instrument. Specifically, the following questions were asked:

- Have you personally experienced any negative/positive effects from bell curve grading in your modules? Please describe your experience.
- Do you think that using the bell curve improves/worsens your university's grading practices? Why?

The responses to both questions reflected an overall more pronounced negative stance towards bell curve grading. Although 27.4% of students experienced positive, or both positive and negative effects from bell curve grading, a higher percentage (55.9%) experienced solely negative effects from it. Similarly,

Table 6. Correlations between fairness of the bell curve grading and motivational constructs by modules

	Modules				
	Language				Content-based
	NUS Compulsory Basic English (n=41)	NUS Elective Ideas and Exposition (n=58)	SNU Compulsory College English (n=24)	SNU Compulsory Advanced English (n=10)	SMU Compulsory Ethics & Social Responsibility (n=78)
Goal Orientation	-.16	.10	.23	.43	.19
A. Approach goal					
Task-approach	-.25	.06	.26	.29	.03
Self-approach	-.19	.07	.41*	-.07	.15
Other-approach	.03	.16	-.22	.65*	.15
B. Avoidance goal					
Task-avoidance	-.24	.01	.27	.21	.19
Self-avoidance	-.23	.08	.51*	.14	.21
Other-avoidance	.01	.03	-.10	.58	.13
Motivation	.02	-.00	.04	-.50	-.09
Interest/Enjoyment	.13	-.08	.00	-.48	-.07
Perceived competence	-.11	-.00	-.13	-.62	-.05
Effort/Importance	-.04	-.09	-.04	.09	.07
Pressure/Tension (inverted)	-.09	.12	.02	-.54	-.23*

Note: *.Correlation is significant at the 0.05 level (2-tailed); **. Correlation is significant at the 0.01 level (2-tailed)

while 34.7% of students believed bell curve grading to improve, or both improve and worsen grading practices, a higher percentage (52.1%) believed bell curve grading only worsens grading practices.

To each question, the responses from every university showed a higher percentage of solely negative responses, compared to relatively smaller percentages of solely positive or a mix of positive and negative responses. In relation to the effects of bell curve grading experienced, 56.9% of the NUS *Ideas and Exposition* cohort and 31.7% of the NUS *Basic English* cohort experienced solely negative effects, as did 53.2% of the SMU cohort and 69.1% of the SNU cohort. This is in contrast to the smaller percentages of those who experienced (i) solely positive effects: NUS *Ideas and Exposition* (10.3%), NUS *Basic English* (7.3%), SMU (30.1%) and SNU (11.8%); or (ii) a mix of positive and negative effects: NUS *Ideas and Exposition* (22.4%), NUS *Basic English* (9.8%), SMU (21.8%) and SNU (8.8%).

In relation to whether bell curve grading improved or worsened the university's grading practices, 51.7% of the NUS *Ideas and Exposition* cohort, 46.3% of the NUS *Basic English* cohort, together with 52.6% of the SMU cohort and 63.2% of the SNU cohort perceived only a worsening of the practices. Again, this is in contrast to the smaller percentages of those who (i) perceived only improvement of the grading practices: NUS *Ideas and Exposition* (13.8%), NUS *Basic English* (14.6%), SMU (35.3%) and SNU (17.6%); or (ii) believed that grading practices would both be improved and worsened: NUS *Ideas and Exposition* (25.9%), NUS *Basic English* (12.2%), SMU (25.6%) and SNU (11.8%).

To further analyse students' conceptions of the fairness of bell curve grading, their responses to the open-ended questions were categorised according to Rodabaugh's (1996) constructs of interactional fairness, outcome fairness and procedural fairness. The analysis surfaced two other student concerns – that of the learning environment, and motivation in learning. There were also comments that could not be categorised according to these themes. With both questions, it is observed that across the universities and the majority of the sampled modules, the three main student concerns were the learning environment, outcome fairness and procedural fairness. Respondents across all three universities made overwhelmingly negative comments about the learning environment. They allude to unwelcome levels of competition and a lack of collaboration as a result of bell curve grading, as well as the spawning of "strategies" to avoid being disadvantaged by the bell curve. Below are some representative comments:

> 159: Due to the bell curve, students are less willing to help other students learn and may hold back on information so that they can achieve an advantage over others. Furthermore, stronger students are known to group together, for group assignment components, so as to ensure that they receive the highest grades available. This prevents those who are not well versed in the subject from getting the most out of the learning opportunity the group assignment is supposed to provide. I've also heard of some of my peers who target classes with weaker students or subjects that are easier to score so that they can improve the grade they receive.

> 135: While it promotes competitiveness among students, I do know of students who try to avoid classes who have the "smarter students" based on speculations so that they get a better shot at getting "A". The whole essence of learning seems to diminish.

Students commenting on "outcome fairness" asserted that bell curve grading did not match their effort or measure their learning:

> 57: Bell curve causes those who have put in sufficiently good effort to still be in the lower grade, just because they did not do as well as the rest. The low grade that students receive does not fully show the amount of knowledge or effort put in by the student, as it is also affected by other students too. Bell curve, in short, does not reflect a person's abilities.

> 103: One or two test questions may decide whether my grade is an A or a B, and I don't think that's necessarily consistent with the knowledge I learned in class or the effort I put into studying.

This was the case even for a student who experienced positive effects of bell curve grading:

> 82: …when everyone is doing not so well, I end up scoring better even though I am not fully aware of what is going on in class.

Comments on "procedural fairness" were not entirely negative. There were students who recognised that bell curve grading could result in "procedural fairness" although some were quick to qualify their comments:

> 107: [t]he bell curve improves grading practices because it's convenient for professors and consistent"

> 148: … in some classes grades are concentrated in the high or low range based on the   difficulty of the papers that year. In one of my classes, the average grade for the finals is consistently a fail grade. I think the bell curve ensures our grades are fairly given compared to previous batches, as difficulty of papers cannot be kept exactly the same."

Overall, the qualitative data showed that among the top three concerns of the student sample pool from the three universities, there was greatest negative sentiment over the learning environment engendered by the use of bell curve grading, although there was also negative perception over its effects on outcome and procedural fairness.

DISCUSSION

In response to the first research question on students' conceptions of the fairness of bell curve grading, the quantitative results reflected that students perceive bell curve grading to be slightly to moderately fair. This appeared to be at odds with the qualitative data which showed a more pronounced negative stance towards bell curve grading, with 55.9% of students experiencing solely negative effects from bell curve grading, mirroring the 52.1% who believed bell curve grading to conclusively worsen grading practices. However, while students from the three institutions, in general, rated themselves highly in relation to goal orientation (m = 5.36) and intrinsic motivation (m = 4.53), they tended to view curve grading fairness less positively (m = 3.43). As observed from the findings, students perceived the use of bell curve grading as having a negative impact on their learning environment, rather than fairness per se, and detrimental to the building of a shared learning culture. This invariably led to various 'coping strategies' that range from avoiding classes with smart students to adopting learning approaches that are highly grade-driven. Moreover, though some students alluded to the importance of bell curve

grading to ensure procedural fairness, they were also concerned about the validity of grading practices on their learning, i.e. in terms of outcome fairness or the notion of the grades as (non) representative of their actual learning.

Given that students were well-aware of the different conceptions of the fairness of bell curve grading, we were surprised that the correlations between goal orientation/motivational levels and perceived fairness were found to be statistically non-significant. In other words, the variation in students' experience with grading practices could not be explained as a basis of the relationship with goal orientation or motivations. Here, we explore alternative explanations to warrant these findings.

First, students are well-aware of the norming nature of bell curve grading; in other words, the assumption of normal-as-average is pervasive and enculturated, what Fendler and Musaffar (2008, p. 82) lamented as "the idea of normal has become formalized and normalized". This is a worrying sign, as we have argued that bell curve thinking if left unquestioned or unrationalized, is detrimental to students' learning experience and emotional well-being (e.g. negative coping behaviours and adopting a fixed mindset). Bell curve grading is entrenched in the psyche of students regardless of goal orientations or motivation.

Second, the students in this study are of high ability and highly motivated, as evident in their self-reports on both goal-orientation and intrinsic motivation. The findings might turn out differently for students with lower academic abilities or motivation. Bell curve grading has inherent structural unfairness (comparing and sorting students based on controlled proportions) which affects all students concerned and certainly warrants more research with different groups of learners.

Third, the nature and quality of the assessment tasks could be taken into account when unpacking the students' perception of fairness of grading practice. Indeed, the investigation of student perceptions of grading practices, or more specifically, the reliability and validity of the grading, would include not only grading practices but also their own understanding of what it means to be awarded a grade. What Brookhart et al. (2016) pointed out in their review was that achievement grades represent a multidimensional measure of success in schools and most teachers' grades were likely to be dependent on both what the students learn and other "academic enablers like effort, ability, improvement, work habits, attention, and participation" (p. 834). They recommend that a study of grading practices should "cast a broader net" (p. 835) to take into consideration antecedent causes, such as instructional and assessment designs, as well as issues such as "working towards clearer criteria, collaborating among teachers, and involving students in the development of grading criteria" (p. 836).

Fourth, there is a need to re-think the conceptualisation of students' understanding of fairness in grading and its impact on other measures of students' achievement motivation. In particular, the argument of involving students in understanding and interpreting grades or grading practices is an important one, and should go beyond grades to using other forms of feedback, such as qualitative comments and even peer feedback. This implies that students could take on the role of assessors and that we should create more opportunities for teachers and students to co-design the assessment and grading processes. In other words, grading is as much about making judgements of the quality of students' work as it is about helping students themselves to develop the capacity for self-evaluative judgement (e.g. Rust et

al., 2003; Boud et al., 2018). Responding to our second research question on how students' goal orientations and levels of motivation relate to their conceptions of the fairness of bell curve grading, the overall mean scores for students' goal orientations and motivations suggest that students were decidedly goal oriented (5.36), and motivated (4.53), and this contrasted significantly with their view of the fairness of bell curve grading (3.43). In terms of goal orientations most students had *task-approach goals* ("do the task correctly" = 5.58) while the smallest percentage were concerned with *other-approach goals* ("do better than others" = 5.12). Correlation between goal orientation and bell curve grading fairness was statistically non-significant, except for certain SNU goals (see Table 6). The *College English* students considered bell curve grading to be fair when they tried to improve on their own grades. However, when they compare their goals with others (external achievement goals), then bell curve grading is considered as unreasonable.

In contrast with the *College English* module, *Advanced English* showed a positive correlation between other-approach goal orientation and bell curve grading (r_s= 0.65, p<0.05) (see Table 6). This indicates that the more students were focused on doing better than others, the more they perceived bell curve grading as a fair measurement. At least two factors need to be considered in this regard: Firstly, the inverted V-curve may be responsible for this conception as it could stimulate competition among students. The qualitative data affirms/illuminates the competition experienced by students. Secondly, the higher English entrance requirement for the *Advanced English* module contributes to more interaction among students in class which may cause students to be acutely aware of their counterparts' abilities and thus highlight their own shortcomings (Hanushek, Kain, Markman, & Rivkin, 2003).

In terms of intrinsic motivation, students showed relatively high levels of motivation at 4.53 as the overall mean, the highest construct at 5.58 was *interest/ enjoyment*. This construct was seen as the largest contributing influence on intrinsic motivation (Tan, 2018, p. 149), suggesting that students were overall motivated in the modules within the survey. Again, when correlated to the bell curve grading fairness, results were statistically not significant, and this was also reflected in the qualitative comments, where motivation was commented upon the least by students. SNU however featured once more with the most marked correlation between grading fairness and motivation. In terms of the correlation between the fairness of bell curve grading and intrinsic motivation (see Table 6), *Advanced English* students showed a moderately negative (r_s=-0.50) correlation. When they perceived bell curve grading as fair, their *interest/enjoyment* (r_s= -0.48), *competence* (r_s= -0.62), and *pressure/tension* (r_s= -0.54) were moderately lower. The moderate tranquility among *Advanced English* students may be partially attributed to the inverted V-curve that generally permits higher grades to a larger population even within a small cohort; however, variables such as teaching methodology, course content, smaller class size, and interpersonal relations may affect students' conceptions as well.

When bell curve grading was perceived to be fair, the data (r=-0.23, p<0.05) showed that SMU students experienced greater pressure and tension. However, the correlation is small and there is an anomaly between the data and the fact that 55 out of the 78 students indicated that bell curve grading was *not* used in the *Ethics* module, with 13 being unsure. For the majority of the students, therefore, the increase in pressure and tension was

probably not attributable to the perceived fairness of the use of bell curve grading in this module. On the whole, these results point to a slight correlation between students' conceptions of bell curve grading fairness and motivation, where decreased motivation and increased pressure and tension are evident when bell curve grading is considered to be fair.

LIMITATIONS

While the use of mixed methods – both qualitative and quantitative self-reports – on students' perceptions allowed us to have a better understanding of how the learning environment of students is affected by norm-referencing grading practices, the findings should be interpreted with caution as the sample size from each institution is relatively small. Future studies should consider replicating this study with larger samples of students, in different institutional contexts and adopting other valid measurement scales on relevant constructs of motivation such as self-efficacy and self-regulation of learning. More research in the future could be directed at the variations in using bell curve grading practices as well as a broader agenda on involving students themselves in the grading process.

CONCLUSION

This study examined students' conceptions of bell curve grading fairness and whether goal orientation and motivation are correlated to their views on fairness. The findings indicated that students have different conceptions of bell curve grading fairness, have formalized view of bell curve grading as a norming tool, and are mostly aware of the inherent structural unfairness. The correlational analyses further highlighted the variation in conceptions of bell curve grading fairness for different groups of learners, given different grading practices and even assess-ment approaches in the modules taught. Two central issues are raised in this article: one, on the need for deeper understanding of the conceptions of students' on bell curve grading fairness and another, on ways to address misconceptions and negative views that hinder their learning. Both issues imply a re-think on how students are provided opportunities to be actively involved in the assessment processes, rather than at the receiving end of assessment outcomes. This re-think will need to include an open and inclusive dialogue by all levels with an interest in enhancing the quality of assessment, teaching and learning in the institution. A possible starting place for this conversation is to debunk the myth that normal-as-average is sacred.

ACKNOWLEDGEMENTS

REFERENCES

Bersin, J. (2014, February 19). The myth of the bell curve: Look for the hyper-performers. Retrieved from http://www.forbes.com/sites/joshbersin/2014/02/19/the-myth-of-the-bell-curve-look-for-the-hyper-performers/#b1d7af313fcf

Biggs, J. B., Tang, C. S., & Society for Research into Higher Education. (2011). *Teaching for quality learning at university: What the student does* (4th ed.). New York; Maidenhead, England; McGraw-Hill/Society for Research into Higher Education/Open University Press.

Bloom, B.S. (1971). Mastery learning. In J.H. Block (Ed.), *Mastery learning: Theory and practice* (pp. 47-63). New York, NY: Holt, Rinehart, & Winston.

Boud, D., Ajjawi, R., Dawson, P., & Tai, J. (2018). *Developing evaluative judgement in higher education: Assessment for knowing and producing quality work*. London: Routledge.

Brookhart, S. (2013). Grading. In J. H. McMillan *SAGE handbook of research on classroom assessment* (pp. 256-271). Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781452218649.n15

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M, T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, *86*(4), 803-848.

Chory, R. M. (2007). Enhancing student perceptions of fairness: The relationship between instructor credibility and classroom justice. *Communication Education*, *56*(1), 89-105.

Close, D. (2009). Fair grades. *Teaching Philosophy*, *32*(4), 361-398.

Czibor, E., Onderstal, S., Sloof, R., & Van Praag, M. (2014). Does relative grading help male students? Evidence from a field experiment in the classroom. Tinbergen Institute Discussion Paper 14-116/V. Available at SSRN: https://ssrn.com/abstract=2488550 or http://dx.doi.org/10.2139/ssrn.2488550

Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of Personality*, *62*, 119-142.

Dubey, P., & Geanakoplos, J. (2010). Grading exams: 100, 99, 98, … Or A, B, C? *Games and Economic Behavior*, *69*(1), 72-94. doi:10.1016/j.geb.2010.02.001

Eichenwald, K. (2012). *Microsoft's lost decade.* New York: Conde Nast Publications, Inc.

Elliot, A. J., Murayama, K., & Pekrun, R. (2011). A 3 × 2 achievement goal model. *Journal of Educational Psychology, 103*(3), 632. doi:10.1037/a0023952

Fendler, L. (2014). Bell curve. In D. Phillips (Ed.), *Encyclopedia of educational theory and philosophy 1*, 84-86.

Thousand Oaks,, CA: SAGE Publications, Inc. doi: 10.4135/9781483346229.n40Ganguly, D. (2015, May 29) Axing the bell curve: Why Microsoft, Cisco did CTRL ALT DEL to this appraisal tool. Retrieved from http://economictimes.indiatimes.com/magazines/corporate-dossier/axing-the-bell-curve-why-microsoft-cisco-did-ctrlaltdel-to-this-appraisal-tool/articleshow/47456433.cms

Grant, A. (2016, September 10). Why we should stop grading students on a curve. *The New York Times*. Retrieved from https://www.nytimes.com/2016/09/11/opinion/sunday/why-we-should-stop-grading-students-on-a-curve.html

Goertzel, T. and Fashing, J. (1981) The myth of the normal curve: A theoretical critique and examination of its role in teaching and research," *Humanity and Society 5*:14-31, reprinted in Readings in Humanist Sociology (General Hall, 1986).

Gordon, M. E. and Fay, C.H. (2010). The effects of grading and teaching practices on students' perceptions of grading fairness. *College Teaching*, *58*(3), 93-98.

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement?. *Journal of applied econometrics*, *18*(5), 527-544.

Hazari, A. (2012, December 2). Why school graders stay away

from the bell curve. *The South China Morning Post*. Retrieved from http://www.scmp.com/lifestyle/family-education/article/1094639/why-school-graders-stay-away-bell-curve

Herrnstein, R. J., & Murray, C. A. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Klapp, A. (2015). Does grading affect educational attainment? A longitudinal study. *Assessment in Education: Principles, Policy & Practice, 22*(3), 302-323.

Kincheloe, J. L., Steinberg, S. R., & Gresson, A. D. (1996). *Measured lies: The bell curve examined* (1st ed.). New York: St. Martin's Press.

Kong, L. (2016). Mimicking religion as coping strategy: The emergence of the bell-curve god in Singapore. *Material Religion*, 12(4), 533-535. doi:10.1080/17432200.2016.1227639

Kulick, G., & Wright, R. (2008). The Impact of Grading on the Curve: A Simulation Analysis. *International Journal for the Scholarship of Teaching and Learning, 2*(2), 1-17.

Lok, B., McNaught, C. & Yong, K. (2016). Criterion-referenced and norm-referencedassessments: compatibility and complementarity, *Assessment & Evaluation in Higher Education, 41*(3), 450-465.

Mcgregor, J. (2013, November 20) For whom the bell curve tolls. Retrieved from https://www.washingtonpost.com/news/on-leadership/wp/2013/11/20/for-whom-the-belcurve-tolls/?utm_term=.59f82a9edf08

Nelson, S. (2010, November 24) Education Reform — Crashing on the Bell Curve http://www.huffingtonpost.com/steve-nelson/education-reform-crashing_b_787303.html

O'Halloran, K. C., & Gordon, M. E. (2014). A synergistic approach to turning the tide of grade inflation. *Higher Education, 68*(6), 1005-1023. doi:10.1007/s10734-014-9758-5

Redding, R. E. (1998). Students' evaluations of teaching fuel grade inflation. *The American Psychologist, 53*(11), 1227-1228. doi:10.1037/0003-066X.53.11.1227 Link

Richert, K. (2018). Why grading on the curve hurts. Retrieved from http://teaching.monster.com/benefits/articles/5658-why-grading-on-the-curve-hurts

Rodabaugh, R. C. (1996). Institutional commitment to fairness in college teaching. *New Directions for Teaching and Learning*, 1996(66), 37-45. doi:10.1002/tl.37219966608

Rust, C., Price, M., & O'Donavan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education, 28*(2), 147-164.

Tan, E. C. (2012, January 20). The bell curve. [Blog post]. Retrieved from https://blog.nus.edu.sg/provost/2012/01/20/the-bell-curve/comment-page-2/

Tan, Y.L.L. (2018). Meaningful gamification and students' motivation: A strategy for scaffolding reading material. *Online Learning, 22*(2), 141-155. doi:10.24059/olj.v22i2.1167

Weil, R.R., and Kroontje, W. (1977). Grade inflation: Causes and cures. *Journal of Agronomy Education, 6*, 29-34.

Ye, L. (2016, November 4). Curve as grade inflation method hurt, not help students. *The Oracle*. Retrieved from https://gunnoracle.com/9563/forum/curve-as-grade-inflation-method-hurt-not-help-students/

Yount, R. (2011, June 30). *The bell curve and assigning grades*. Retrieved from http://www.drrickyount.com/2011/06/jason-norris-and-the-bell-curve/