

## Introduction

As with many other tertiary level education processes and institutions, the assessment and marking process in higher educational institutions (HEI) has come under increased scrutiny from an instrumental perspective of education (Olsen, 2007; Atwood, 2008). If as this view proposes, education's role is to be an engine of national economic growth then it needs to offer sufficient and appropriate skills, knowledge and competences to address an assumed convergence between the goals of individuals and those of a post-industrial economy and society (Johnston & Watson, 2004; Williams, 2008). From this perspective, educational aims are focused upon competence acquisition by individuals.

However, this view can be in conflict with the aim of the 'marking process'. As Sadler (2009:807) states "The integrity of grades has to do with their authenticity and provenance rather than their utility for various purposes...". There is potential for significant disquiet with those tasked to grade assessed submissions as they seek to reconcile these two competing perspectives. As a result, in the UK a common normative metric of competency achievement is found in the award of 'good quality degree' marks, which are interpreted as comprising the proportional allocation of first and upper second class awards to work presented for undergraduate assessment or merit/distinction for postgraduate assessment (even though the determination of this volume is typically not undertaken by *grading proportion* in the UK (Sadler, 2009)). When these are then included as part of comparative HEI measurements to generate ranking tables, there is inevitably a tension within the marking process to ensure that it remains both objectively and subjectively understood to be fair and robust yet contributes to the external confirmation of teaching and assessment quality. However, this view risks ignoring a range of other potential variables that may shape the outcome of the marking process. These variables can for example reflect the cognitive process of marking, learner diversity and the situational context of the learning (Illeris, 2012).

This paper develops a conceptual framework of the marking process by identifying potential influences that can combine in varying ways to generate an extrinsic measurement of achievement. We seek to understand this process in that context. It is one of the four threats to grade integrity identified by Sadler (2009). Through data gathered by HEI markers whilst marking submitted assessments, an analysis of that data is undertaken to identify the potential and comparative importance of those influences. This will help clarify understanding about the marking process and potentially instrumental influences. It contributes to both the ongoing debate about comparative institutional performances and grade inflation, as well as offering developmental opportunities for academic staff. We finally suggest areas for further research to build upon the emergent outcomes and opportunities of this work.

## The intention to grade

We firstly adopt a normative stance that HEI markers aim to undertake objective grading of submitted assessments. In other words, it is their intention to practice academic grade integrity. This is a very important assumption as it underpins a large scope of subsequent sector, policy and life decision outcomes that are built upon the premise of grade integrity (Sadler, 2009). For reasons outlined in subsequent sections of this paper however, actual behaviour may not deliver upon this intention, but intention is a good proximal measure of expected behaviour (Francis et al, 2004). An intention to act is arguably determined by the attitude of the actor, the social pressure upon the actor and the actor's belief in how much control they have and the ease with which that intention can be delivered (Ajzen, 1991). We seek to determine the relative importance of these potentially influencing variables upon the intention to deliver an objective grade. Thus we undertake a principal component analysis (PCA) to seek to reduce the number of potential influencing variables to those which have significant impact upon the grading intention. Whilst behaviour is observable, intention is a cognitive construct. Intention is not directly measurable. Instead we next identify appropriate manifest variables that could illuminate a latent intention.

For many students, an understanding of the UK HEI grading system can be something new and complex. HEI grading can be (although is not exclusively) based upon a class based system that reflects achievement with an attainment of evidence that addresses identified intended learning outcomes that are (typically) articulated through assessment criteria. It has become increasingly common practice that UK HEIs are encouraged to adopt the full marking range (from 0 to 100%) including the arts and humanities disciplines, that have historically been presented as operating a narrower marking range (Barnes *et al*, 2001). Sadler (2009:809) presents a robust view of the normative factors that contribute towards establishing a grade for an assessed submission. In this work we are concerned with the reality of that process, particularly items 1,2 and 3, where the individual marker seeks to resolve and balance *in real time* influences that give rise to an allocated grade.

“(1) Students deserve to have their work graded strictly according to its quality, without their responses on the same or similar tasks being compared with those of other students in their group, and without regard to the students’ individual histories of previous achievement.

(2) Students deserve to know the bases on which judgments are made about the quality of their work. There should be few if any surprises.

(3) Students deserve their grades to have comparable value across courses in the academic program in which they enrol, and across the institution. Courses should not exhibit characteristically tough or lenient grading.

(4) Students deserve grades that are broadly comparable across institutions and maintain value over time, so that the standing of their educational qualifications is protected not only by the college or university in which they study but also by higher education as a social institution”

Allen (2005), Sadler (2009) and Bloxham and Boyd (2011) have argued that grading can generate outcome marks which do not seem to follow a given marking methodology or practice, as markers seek to justify an awarded mark and reward those students who live up to their expectation of what a ‘good’ quality output is (see (2) above). There are tensions in grade allocation which have been labelled as cultural resistance to absolutes in an arts context for example, where absolutism conflicts with the inherent relativism of the arts (Brown, 2004). As a general guide, assessment and grade determination has moved towards being criterion based and away from a subjective sense of student achievement (*gatekeeping* or *norm assessment*).

Sadler (2009) identifies heuristics as an influence upon grade integrity but from a post grading adjusting process in order to reflect a perception of cohort quality, rather than during the grading process. We focus upon the grading process. Such heuristics could include the need for more distinctive differentiation between the classification thresholds which can be skewed by a *symbolic comparison bias (SBHE)* (i.e. it is easier to allocate a 62% when compared with a 67% or a 68%). Crisp (2008) has also identified that markers may also evidence other heuristics during the assessment process; the *affect heuristic (AFH1)* is an emotional response to a submission and the *availability heuristic (AVHE)* can influence grade allocation as the information in an assessed response is not located where it is expected but found elsewhere. Other potential heuristics from the literature can include ego depletion (*EGDP*) where repeated attention given to a single task, weakens an individual’s decision making capacity (Kahneman, 2011) and which might give rise to an influence from when a given script is graded – whether it is near the start of the grading period (*POW0*) or the middle (*POW1*) or the end (*POW2*) of the marking period.

The flexibility of the marker towards the rigorous use of the academic register (*ACRE*) has also been identified by Cadman (1997) as potentially impacting upon the grade allocation. Grade clustering (which can arise for a number of reasons such as historical patterns, previous practice and the more discursive and subjective nature of the arts and humanities disciplines) is also a potential grade influence (*DIDI*) although this may be mitigated by the use of criterion based marking (*ACRU*) (Brown, 2004). In a similar vein, the disaggregation of marks by multiple small overall assessment tasks can further exacerbate grade clustering (*ASDI*) (Brown, 2004).

Brown (2004) further suggests that the third class degree is disappearing in UK HEIs and markers may be institutionally influenced to reduce the use of this classification range (*COPI*). Finally, from both a review of available literature, considerations with colleagues and the variables that could arise from an examination of

learning styles and the situation context of learning (Illeris, 2012), other potential variables were identified with colleagues that may impact upon the marking process are proposed as:

- Previous formative comments upon drafts (*FORM*).
- The measured 'quality' of incoming students to programmes of study (as determined by input tariff) (*INPR*)
- The volume of contact time with a student pre assessment (*VOCO*)
- The scope of contact time pre assessment (*SCCO*)
- The type and validity of second marking / moderating practices (*SMMP*)
- The extent of supportive training and development for staff of the marking process (*INTR*)
- The actual length and duration of the process of assessment grading for a staff member (*APST*)
- The allocation of a grade from a tutor incorporating developmental aims (e.g. by awarding a generous mark or similarly through awarding a low mark) (*DEAI*)
- The determination of a grade may be influenced by the volume of previously awarded 'quality' grades (>60%) for example (*ALRA*)
- The grade allocated to a completed script which follows on from a poorly graded script (*POR1*), a highly graded script (*POR2*) or an averagely graded script (*POR3*)
- The historic development of the marker may have established disparate patterns of behaviour (*CUPR*) (Hand & Clewes, 2000: Allen, 2005).

Crisp (2008) has also explored some of the psychology of marking for University entry examinations and noted that some (but not all) markers were observed to undertake a process of 'self-control focus', whereby they consider the situation of writing and the writer, so as to establish a personal view of the qualities of the work being considered, but which may introduce differential grading practices. Some markers therefore did – to an extent – try to reconstruct the intended meaning of a submitted assessment task and saw that as an important part of the assessment process (*ANCO*). The extent of accommodation of variations in language competence was argued to be dependent primarily upon the marker and the discourse (knowledge base) they belonged to (Crisp, 2008).

Available evidence also shows wide variation in mark achievements and methods of assessment by faculty and by institution (Yorke *et al*, 2000: 2002). Barnes *et al* (2001) called this *adaptation level theory* where in different academic disciplines, students (and staff) that are perceived as less academically competitive migrate to those degrees that are perceived as less demanding in order to secure higher exit awards (*PEDE*).

All of the potential variables that are focused upon during the decision making stage of grading reflect (1) and (2) of Sadler's concept of grade integrity. A number of the variables identified may influence the final grade allocation but are not captured at the point of initial decision making (see **table 1**). The next section outlines the methodological approach to the design of the research and the construction of the data collection tool that addresses the manifest variables used to reflect understanding about the factors identified in the literature.

## Methodology

This paper adopts the arguments and framework from the Theory of Planned Behaviour (TPB) and considers the relationships between the markers, their pressures, control and attitudes towards the intention to allocate an objective grade (Ajzen, 1991). We wish to determine the relative importance of identified manifest variables upon the intention to allocate a grade *during* the grading process. Hence, an intention to allocate a grade is determined by the marker's attitude and the prevailing subjective norms, behavioural and normative beliefs they recognise and value. Intention is then influenced by this 'combined' function of beliefs and norms. Attitude is a multidimensional factor shaping individual intentions and is described as a belief about how likely a given consequence of individual action is, plus whether that consequence is perceived to be good / bad by the individual. The subjective norm of an individual is the outcome of normative beliefs (in terms of what their referents would do in that same situation), moderated by the desire of the individual to conform to that referent's predicted actions.

TPB acknowledges that there can be instances of a desire to act, but then an inability to do so by the individual. Unpacking the PBC concept identifies perspectives of perceived control (PC) over a task and perceived difficulty (PD) of undertaking the task (Kraft et al, 2005). The TPB framework therefore argues that an intention to perform a behaviour is determined by the following latent variables:

- (1) subjective norms (SN) of the individual which can be argued to be determined by a) the group norms of marking expectations (GN), b) the views of significant others and associated issues (SO) and c) the social identity (SI) the individual derives from participation in the process and their standing in the academic community
- (2) the emotional attitude and engagement towards the behaviour (AE)
- (3) the perceived behavioural control (PBC) in the ability to perform the behaviour (perceived difficulty (PD)) and their opportunity to perform the behaviour (perceived control (PC)).

We will analyse the collected manifest data using Principal Component Analysis (PCA) to determine the correlations between those variables and the resultant latent constructs with the aim of reducing them to generate a smaller prioritized selection of clustered factors that reveal influential variance. SPSS (v23) is used to evaluate the collected data.

In the table below the variables are presented and grouped. This paper is focused upon that period of time and cognition which generates an allocated grade to a submission and which therefore includes only those manifest variables relevant to that stage (2) in the marking process. **Table 1** presents the conceptual relationships between the manifest variables and the factor headings presented for the TPB.

**Table 1:** About here

### *Development of the data collection tool*

The application of PCA requires a significantly sized sample to be gathered of manifest variable data to ensure sufficient overall validity and support reliability measures of the determined latent constructs. It is also desirable practically to seek to engage a range of assessment activities. Initial aims to access wider cross institutional data collection met with limited enthusiasm (the pressures of a high marking workload was a major factor inhibiting participation), hence the sample is smaller than intended. Nonetheless, convenience undergraduate and postgraduate samples of anonymised work, were available from 2 HEIs in the North West of England. Grading policies were reviewed to ensure they were broadly similar across the units/modules being graded (i.e. by aggregate scores, stated ILOs and constructed assessment criteria (Sadler, 2009)):

**Table 2:** About here

The development of the data collection tool, in keeping with the accepted techniques typically used in TPB applications employs a self-administered questionnaire (Francis *et al.*, 2004; Blunch, 2008). The data collection commenced at the start of Semester 2 (2016-2017), using a three staged data capture tool. This was undertaken both pre, during and post the marking process by the participants. There are no assumed (or argued for) expected co-variances between the errors for each manifest variable in the model.

In order to reflect the dynamic of the marking process, data capture needed to be undertaken *at the point of* grade decision making. This is a challenging requirement for data capture, particularly when time is constrained. For this reason, we viewed data collection through three activity stages:

- 1) Factual Information relating to pre - assessment
- 2) Cognitive Information relating to the actual assessment and grade decision
- 3) Reflective Information relating to grade review and moderation of the assessment and cohort

As this paper is focused upon (2) above, it was a priority to ensure that data collection should not be burdensome or disproportionate to the work commitment to undertake the marking by participants. We also rejected reflective methods that rely upon a post grading discussion of influential factors. This was to mitigate

the influence of memory paradigm, where recollections of reasons for a behaviour may shift after the completion of that behaviour (the marking process) to be closer to what a marker believes would be objective practice (Fischhoff and Beyth, 1975 cited by Kim, 2017). Proposed data collection methods considered therefore included:

- 1) Thinking aloud data capture through a short commentary on each assessment marking (Ljungberg et al, 2013). Whilst this data could be recorded for later coding, the potential volume of coding required especially if undertaken by a single individual to ensure consistency, is significant.
- 2) Development of a tool to allow a visual capture of the data that is fast to complete yet is sufficiently detailed to capture the views to the key questions. This includes the translation of the Likert scales for specific questions into visual representations that can be quickly commented upon by the participant. Onwuegbuzie et al (2010) offered scope and discussion to construct a visual material capture tool.

During preliminary conversations with colleagues, option (2) was chosen with an e-questionnaire, mainly because of convenience. An excel workbook was therefore developed to capture data which through some visual basic programming, allowed repeated data entry for successive marked scripts. Data entered at that time, was subsequently stored and compiled to a hidden worksheet which was then used for analysis. Markers used a mix of Grademark via Turnitin and paper based marking and inevitably a small extra time burden was incurred capturing the decision making data. Pilot feedback suggested this was approx. 30 seconds/graded script. The questions and their construction are discussed shortly.

Francis *et al.* (2004) note that for TPB questionnaires, questions regarding a particular intention can be asked directly or indirectly depending upon the confidence that can be exhibited by the marker in giving a manifest judgement regarding a latent variable.

- Direct questions can be asked where there is high confidence that the marker is able to pass a judgement on the value of a given intention. For example - *The assessed work is (not) within the set word count limit and does (not) therefore incur an automatic penalty.*
- Indirect questions are appropriate where it cannot be assumed that the marker will be able to pass an authoritative judgement on the outcome of a given grade allocation. For example – *The grade I allocate will result in a positive / negative stimulus for the student concerned.*

We have assumed markers will be able to comment directly on all identified manifest variables. Data capture of these beliefs and norms generally uses simple Likert based measurements of attitude and how easy / difficult a behaviour would be to enact. This paper adopts the decimal scale promoted by Brody and Dietz (1997), Shulruf *et al.* (2008) and Carifio and Perla (2008), which helps to mitigate potential bias arising from the value perceptions in the standard application of 1-5 or 1-7 Likert scale.

In terms of collecting and scaling the responses for direct (and indirect) questions, the likelihood of a marker's action is scaled by the outcome's desirability. Care must be taken in how this multiplication is achieved so as to correctly interpret the data.

- Francis *et al.* (2004) advocate that for the measurement of attitudes towards an intention, we need to calculate attitudinal antecedents towards an issue on a unipolar Likert scale (say +1 to +10), but for outcome evaluations (which can be viewed as negative, neutral or positive) we should adopt a bipolar Likert scale (say -3 to +3). **Tables 3 and 4** summarises these scales and weights by latent construct.

**Table 3:** About here

**Table 4:** About here

## PCA Analysis of the Marking Data

157 data responses is acceptable for undertaking PCA sampling (Kline, 2008). After data capture, manifest sub-construct tables were then compiled as required. These formed the basis of the PCA extraction with SPSS. PCA is recommended for behavioural research and follows from the narrative of Sparks (2007). Any item not correlating on any other item (at  $<0.4$ ) was ignored in the analysis. The parameters of the initial analysis were: Rotated Factor Solution (Varimax), Eigenvalues  $>1$ , and convergence limited to 25 iterations. Bartlett's test of sphericity was significant ( $p<0.000$ ) and the Keyser-Meyer-Olkin measure of sampling adequacy was 0.559, exceeding the minimum standard required (0.5) (Field, 2013). Reliability measures (Cronbach's Alpha) are provided for the latent constructs identified. The final solution identified 5 components that explained a high 80% of observed variance shaping the allocation of a mark.

**Table 5:** About here

**Table 6 :** About here

**Table 7 :** About here

**Component 1** – loads heavily on EGDP and FORM, as well as AVHE and ANCO. This cluster highlights the marker's individual control and attitude towards the work. We might describe this cluster with the label of “**Active Review**”, in that the full scope of the work and its meaning is sought and considered in the same manner as the preceding script. Whilst the mean of EGDP was 3.45 indicating a generally consistent allocation of time per script although there was also significant variation in this view (with a standard deviation of participants views of 3.11). Prior engagement with the tutor to receive formative feedback is also stressed.

**Component 2** – loads heavily on ACRU, AFH1, AVHE and then ANCO. This cluster highlights the marker's individual control and attitude towards the work through active use of assessment criteria for all the material in a submission of personal interest to the marker. We might describe this cluster with the label of “**Interpretive Alignment**”.

**Component 3** – loads heavily on SBHE and ALRA only. Whilst the reliability of this component construct is low, the combination of these two variables suggests markers have some difficulty in determining a grade for a script when it follows a similar preceding script and there is a consideration of the overall number of already allocated quality marks ( $>60\%$ ). We might label this cluster as “**Similarity Quality**”.

**Components 4 and 5** – load only on single variables of POW1 and POW2 respectively. Whilst reliability measures are not therefore applicable, these components suggest that the position of the script in the second half of a marking period are active inputs shaping the intention of allocating a mark.

## Discussion

The scope of potential variables shaping the decision to allocate a final grade potentially occurs at both pre and post that initial grading decision. Of the other variables concerned with the decision making point of grading, the normative assumption made of markers seeking objective grading of assessments has been supported in the identified factor **components 1 and 2 (Active Review and Interpretive Alignment)**. These manifested as recurrent use of assessment criteria, comparative allocations of time per script and using all available information in the presented script to support the construction of a grade. This is a positive affirmation of Grade Integrity reflecting the first and second items of Sadler's (2009) definition. These components also stressed the input of personal interest by the marker and previous formative feedback already presented on earlier versions of the student work (Crisp, 2008).

Factor components 3, 4 and 5 – are more problematic in both their reliability within the analysis but also interesting in the suggestion of other sources of influence upon the grading decision. In terms of the latter, the findings suggest that the allocation of a grade is exacerbated by script similarities and the overall recalled view of the number of quality marks already allocated (Kahneman, 2011). This could apply to an extent to a number of high/low/average perceived quality scripts being marked in sequence, although components 4 and 5 also suggest the position of the script in the second half of a batch of marking is a consideration. Thus these components suggest some input of expectation conformity in what the grades allocated would be, given the preceding perception of the marker on the overall number of quality grades already allocated, the immediate preceding script grade and position of the script in the marking period.

Interestingly, no evidence emerged during the cognitive stage of grade allocation that reflected the determined quality of the preceding script (*POR1,2 or 3*). Furthermore, from the discussed literature, cognitive heuristics emerged as the most important in shaping that grade allocation, whilst the potential influences derived from available marking time (*VOCO, APST*), previous marker history (*INPR, COPR, PEDE*) and the context of the marker (*ASDI, COPI, INPR*) did not emerge as influential at the point of making the cognitive decision of a grade.

## Developmental Implications

This work was undertaken within departments that deliver similar programmes from the same epistemological discipline (*DIDI*). We might expect wider variations if we were to compare data from say the sciences with this humanities outcome. However, we believe there is scope for using the framework and data collection tool to be employed as a pedagogical template to help review and as required, identify areas for development within the marking process. Initial recommendations from this first report of work might be:

- Break up marking into smaller periods of activity without external distraction
- Avoid compiling cohort profile data until the marking process is complete for the cohort
- At the institutional level, ensure sufficient time is allocated to ensure the full scope of a submission can be considered by the marker
- Ensure assessment criteria are robust and clear for all markers concerned

We acknowledge that a number of discussed manifest variables were not directly considered in this analysis which was focused upon the cognitive stage of allocating a decision. Some of these omissions were also for research design reasons. For example, all the assessed work was undertaken anonymously and therefore the scope for using a grade developmentally is discounted (*DEAI*), the marker of an assessment may or may not be involved in the delivery of the unit's materials and thus delivery issues were discounted (*VOCO, SCCO*) and all the grade allocations made were first marks prior to any form of moderation (thus discounting *SMMP*). *PEDE* is difficult to capture without a follow up activity with the markers and in particular the students concerned, which was not feasible for practical reasons. However, it is our intention to explore those variables in subsequent work drawing upon this data and outcomes

This work also illuminates issues in the recurrent concern of grade inflation. Sadler (2009: 823) is clear that the important focus here should not be upon the grade but upon the alignment of the grade with the level of achievement. Markers could be expected to therefore construct their assessment criteria appropriate to accepted external standards (although the degree of customisation by the marker for an assessment is open to debate) and undertake frequent reflection upon them during the marking process. This was the core finding of this work. However, this work does not comment upon the construction of the assessment criteria per se, but **Component 2** included a heavy weighting upon the active use of assessment criteria (ACRU) from the markers. This lends partial support that if grade inflation was perceived to be occurring, at least for the sample concerned, it is more likely to be a concern with the assessment criteria construction than with the decision making during grading. We should not therefore conflate the marking process with grade inflation. Our work aligns with Sadler (2009) that these are different processes with different developmental implications.

Whilst this analysis opens a window onto the cognition of the grading decision, it also leaves a large number of questions unresolved and to be explored further. A larger sample with additional data on the actual grade allocated, reflective views of the overall marking experience (in a given period for a given cohort) and

assessment task construction information – may further illuminate the full marking process with for example the exploration of relationships between those manifest variables and the actual grade allocated to a script, in other words, further work is required on the integration of the three outlined stages of the marking process presented earlier.

Securing access to data that reflects how markers reach grade allocation decisions was not easy or without resource implications. We acknowledge that and are grateful for those who participated in this work. We believe that this approach is scalable, offering further insights into this process. However, the emergent outcomes from this first focus suggest that participants sought to maintain a normative view of the grading decision, is reassuring although there was the potential for that position to be influenced by a number of other biases.

## References

- Ajzen, I., (1991), "The Theory of Planned Behaviour", *Organizational Behavior and Human Decision Processes*, Volume 50, Issue 2, pp. 179-211.
- Allen J.D., (2005), "Grades as valid measures of academic achievement of classroom learning", *The Clearing House*, Volume 78, Issue 5, pp. 218–223.
- Atwood, R. (2008), "Love of learning lost in 'studying for jobs' ", *Times Higher Education*, available at <http://www.timeshighereducation.co.uk/story.asp?storycode=400676> (Accessed 14<sup>th</sup> January 2017).
- Bagues,, M., Labini, M.S. and Zinovyeva, N., (2008), "Differential Grading Standards and University Funding: Evidence from Italy", *CESifo Economic Studies*, Volume 54, Number 2, pp. 149–176.
- Barnes, L.B., Bull, K.S., Campbell, N.J. and Perry, K.M., (2001), "Effects of academic discipline and teaching goals in predicting grading beliefs among undergraduate teaching faculty", *Research in Higher Education*, Volume 42, Issue 2, pp. 455–467.
- Black, D. and Williams, P., (2009), "Inside the Black Box: Raising standards through classroom assessment", *Phi Delta Kappa*, available at <http://www.rdc.udel.edu/wp-content/uploads/2015/04/InsideBlackBox.pdf> (Accessed January 20<sup>th</sup> 2017)
- Bloxham, S. and Boyd, P. (2011), "Mark my words: the role of assessment criteria in UK higher education grading practices", *Studies in Higher Education*, Volume 36, Issue 6, pp. 655–670.
- Blunch, N.,(2008), *Introduction to Structural Equation Modelling Using SPSS and AMOS*. Sage, London.
- Brody, J. and Dietz J., (1997), "On the dimensionality of two question format Likert Attitude scales", *Social Science Research*, Volume 26, Issue 2, pp. 197-204
- Brown, R., (2004), Out of range, available at <http://78.158.56.101/archive/palatine/files/1349.pdf> (Accessed 3<sup>rd</sup> April 2014).
- Cadman, K.(1997), "Thesis writing for international students: A question of identity?", *English for Specific Purposes*, Volume 16, Issues 1, pp. 3-14

- Carifio, J. and Perla, R.,(2008), “Resolving the 50-year debate around using and misusing Likert scales”, *Medical Education*, Volume 42, Issue 12, pp. 1150-1152
- Crisp, V., (2008), “Exploring the nature of examiner thinking during the process of examination marking”, *Cambridge Journal of Education*, Volume 38, Issue 3, pp. 247–264.
- Field, A. (2013) *Discovering Statistics using IBM SPSS Statistics* (fourth edition), Sage, London.
- Francis J.J., Johnston M., Eccles M.P., Grimshaw J. and Kaner, E.F.S., (2004), “Measurement issues in the theory of planned behaviour: A supplement to the Manual for constructing questionnaires based on the Theory of Planned Behaviour”, available at <http://www.rebeqi.org/ViewFile.aspx?itemID=212> (Accessed 10<sup>th</sup> September 2012).
- Hand, L. and Clewes, D., (2000), “Marking the difference: an investigation of the criteria used for assessing undergraduate dissertations in a business school”, *Assessment and Evaluation in Higher Education*, Volume 25, Issue 1, pp. 5–21.
- Illeris, K., (2012), A comprehensive understanding of human learning, Illeris K (Ed.), *Contemporary Theories of Learning*, Routledge, London, pp.7-21
- Johnston, B., Watson, A.,(2004), “Participation, reflection and integration for business and lifelong for business and lifelong learning: Pedagogical challenges of the integrative studies programme at the University of Strathclyde Business School”, *Journal of Workplace Learning*, Volume 16, Issue 1-2, pp. 53-63.
- Kahneman, D., (2011), *Thinking Fast, Thinking Slow*, Penguin, London.
- Kim, N.S., (2018), *Judgement and decision-making in the lab and the world*, Palgrave: London.
- Kline, P.,(2008), *An Easy Guide to Factor Analysis*, Routledge, Hove.
- Kraft, P.R., Jostein S. and Stephen-Røysamb, E.,(2005), “Perceived difficulty in the theory of planned behaviour: Perceived behavioural control or affective attitude?”, *British Journal of Social Psychology*, Volume 44, Issue 3, pp.479-496
- Ljungberg, M.K., Douglas, E., McNeill, N., Therriault, D. and Malcom, Z., (2013), “Re-conceptualizing and de-centering think-aloud methodology in qualitative research”, *Qualitative Research*, doi: 10.1177/1468794112455040 (Accessed 3rd February 2017)
- Olsen, J.P., (2007), The Institutional Dynamics of the European University, Maassen P. and Olsen, J.P. (Eds.), *University Dynamics and European Integration*, Springer, Dordrecht, pp.25-54.
- Onwuegbuzie, A. J., Leech, N. L., & Collins, K. M., (2010),”Innovative Data Collection Strategies in Qualitative Research”, *The Qualitative Report*, Volume 15, Issue 3, pp. 696-726.
- Sadler, D.R., (2009), “Grade integrity and the representation of academic achievement”, *Studies in Higher Education*, Volume. 34, Issue 7, pp. 807–826

Sparks, B. (2007), "Planning a wine tourism vacation? Factors that help predict tourist behavioural intentions", *Tourism Management*, 28 (5) (2007), pp. 1180-1192

Shulruf, B., Hattie J. and Dixon, R.,(2008), Factors affecting responses to Likert Type questionnaires: Introduction of the ImpExp, a new comprehensive model, *Soc. Psychol Educ*, Volume 11, Issue 1, pp.59-78

Williams J.,(2008),"Constructing social inclusion through further education – the dangers of instrumentalism", *Journal of Further and Higher Education*, Volume 32,Issue 2,pp.151-160

Yorke, M.,(2002), "Subject Benchmarking and the assessment of student learning", *Quality Assurance in Education*, Volume 10, Issue 3, pp.155-171

Yorke, M., Bridges M. and Woolf, H.,(2000), "Mark distributions and marking practices in UK higher education: some challenging issues", *Active Learning in Higher Education*, Volume 1, Issue , pp. 7-27