Improving assessment tasks through addressing our unconscious limits to change

Gordon Joughin, Phillip Dawson and David Boud[1]

Centre for Research in Assessment and Digital Learning, Deakin University, Geelong, Australia;

[1] also, University of Technology Sydney and Middlesex University, London

Abstract

Despite widespread recognition of the need to improve assessment in higher education, assessment tasks in individual courses are too often dominated by conventional methods. While changing assessment depends on many factors, improvements to assessment ultimately depend on the decisions and actions of individual educators. This paper considers research within the 'heuristics and biases' tradition in the field of decision making and judgement which has identified unconscious factors with the potential to limit capacity for such change. The paper focuses on issues that may compromise the process of improving assessment by supporting a reluctance to change existing tasks, by limiting the time allocated to develop alternative assessment tasks, by underestimating the degree of change needed, or by an unwarranted overconfidence in assessment design decisions. The paper proposes countering these unconscious limitations to change by requiring justification for changing, or not changing, assessment tasks and by informal and formal peer review of assessment task design. Finally, an agenda for research on heuristics and biases in assessment design is suggested in order to establish their presence and help counter their influence.

Key words: assessment design; change; heuristics; biases; decision-making

Introduction

Why is it so hard to change assessment? Despite decades of research and theorising on assessment and how it can be improved, traditional forms of assessment continue to dominate in much of higher education. Changing assessment practices is a complex business that extends well beyond the decisions of individual academics to include a wide range of factors in the broader disciplinary and institutional environment. Nevertheless, the decisions of individual academics and teaching teams in planning assessment tasks for their courses remain central to improvement. Without good buy-in from those who will implement change within courses and modules, innovation prompted by macro changes will be severely inhibited. While much good practice occurs at the course/module level, the focus of this paper is on a set of factors which inhibit change at this level, which have not been previously explored in this context, and which need to be taken into account in any systematic effort to improve assessment within individual courses or programs.

This paper commences with a summary of the now well-rehearsed arguments for changing assessment and previously noted impediments to change. In order to better understand the limits to change, the paper then focuses on contrasting approaches to decision making that highlight the role of unconscious influences which can distort the deliberative processes required by sound assessment design, outlines a number of specific factors identified by decision making research that may be particularly influential in inhibiting change, suggests ways of ameliorating this influence and, finally, proposes research to improve understanding and practice in relation to this aspect of assessment design.

The challenge of assessment change

The case for reforming assessment and suggestions about how this should be done have been made with particular cogency over the past decade, building on a long tradition recognising the centrality of assessment to student learning and acknowledging the need for assessment to provide trustworthy evidence of student achievement. An abundance of scholarly literature supports the case (see, for example, Boud and Falchikov 2007; Joughin 2009; Merry, Price, Carless, and Taras 2013; Kreber, Anderson, Entwistle, and McArthur 2014). At the same time, 'manifestos' for change have succinctly captured expert opinion on best practice in the UK (Price, O'Donovan, Rust, and Carroll 2008) and Australia (Boud and Associates 2010) and government imperatives for assessment have been expressed in official documentation in Australia (Commonwealth of Australia 2015), the UK (The Quality Assurance Agency for Higher Education 2014), the USA (Ewell 2013) and elsewhere. The consensus from these sources is clear: amongst its multiple functions, assessment needs to provide evidence of integrated learning at the program level, support learning through assessment tasks that are learning processes in themselves, develop student autonomy in judging their own work, and prepare students for the world of practice beyond graduation. To perform these functions, significant change in assessment methods is essential and to effect these changes, academics' decision making in designing assessment tasks and associated practices plays a central role.

If change is needed, change processes have received surprisingly limited consideration in assessment literature. A small body of work draws on change theory in relation to the broad field of teaching and learning in higher education (see, for example, Trowler, Saunders and Knight 2003; Bromage 2006) and this work has informed more specific studies. For example, a recent study of attempts to change assessment 'from above' in response to government initiatives identified 'patterns of resistance' from academics (Deneen and Boud 2014). A consideration of changing assessment in the context of higher education institutions as complex adaptive systems has argued that attention should be given to several organisational levels. These include the institution as a whole, the department, the degree program, and the individual course, with different matters requiring attention at these different levels. Within this framework, factors likely to have a strong influence on assessment include institutional assessment policies, departmental culture, common practices within disciplines, and academics' experience of their teaching environment. In relation to the latter, academics' workload, their knowledge of assessment issues and options, and how their efforts to improve assessment are recognised and rewarded are all likely to impact strongly on their engagement with improving assessment tasks (Macdonald and Joughin 2009).

Some have argued for the primacy of the departmental or workgroup 'meso' level in effecting change, while acknowledging the 'macro' level of the organisation as a whole and the 'micro' level of the individual academic as change agent (Trowler, Fanghanel, and Wareham 2005). The meso level is often seen as the critical site of engagement between students and teachers where 'changes actually take place' and that, consequently, 'a good theoretical understanding of how and why this happens – or doesn't happen – is therefore necessary' (Trowler, Fanghanel, and Wareham 2005, 435). According to this perspective, the assumption of 'methodological individualism', that is, that assessment change can be driven

at the 'micro' level of the individual teacher is a common but mistaken understanding of how change occurs in organisations.

Substantial improvements to assessment practice undoubtedly require action at various levels and in various spheres of organisational activity. Moreover, the different levels and factors are in dynamic interaction with each other. Within this complexity, however, recent research has drawn attention to the significant role of individual academics and teaching teams in designing assessment tasks and associated practices at the micro level of the individual course (Bearman et al. 2016; Boud et al. 2016). These studies have established that, while operating within the macro environment of the institution and the more immediate meso environment of the discipline or department, individuals and teams do make significant decisions about the nature of the assessment tasks in their courses, including whether to change them or not and the nature and extent of change where this is seen to be needed. At this level important improvements to assessment tasks can occur, so that it constitutes another level where 'changes actually take place' and hence one where 'a good theoretical understanding' of the dynamics of change, including approaches to decision making that limit the scope of change, is necessary.

Contrasting approaches to decision making

We tend to think of decision making as a conscious, rational process. However, many decisions in everyday life and in academia, including decisions regarding assessment tasks, are made without conscious thought – they are more-or-less spontaneous, automatic, quickly made and often based on extensive relevant experience. Such decisions are essential to personal and organisational functioning and usually serve us well. However, research over the past forty years has demonstrated our propensity to make quick decisions that deviate from decisions we would make if we adopted a more systematic, deliberate approach. These two approaches – the rapid and automatic on the one hand, and the more thoughtful, deliberative on the other – have been labelled as System 1 and System 2 thinking in both scholarly and popular literature on decision making (Stanovich and West 2000; Kahneman 2011). The distinction between them is central to this paper so a short explanation of each is in order.

The terms System 1 and System 2 were coined to summarise a range of 'dual process theories' of reasoning in judgement and decision making. System 1 'is characterized as automatic, largely unconscious, and relatively undemanding of computational capacity' while System 2 employs controlled, analytical processes (Stanovich and West 2000, 658-659). System 1 has been typified as 'fast' and includes both the intuitive and usually accurate thinking of the expert and the potentially distorted thinking of the non-expert. System 2 in contrast involves mental work that is 'deliberate, effortful, and orderly', able to 'construct thoughts in an orderly series of steps', and, under the right circumstances, able to take over from the 'freewheeling' System 1 (Kahneman 2011, 20–21). Assessment theory and recommendations for improving assessment invariably, and understandably, are based on a conscious, analytical, System 2 approach. Academics grappling with re-designing assessment tasks for their courses may engage both systems at different points, conscious of their analytical processes but unaware of the influence of System 1 which, as Kahneman concludes, 'is more influential than your experience tells you, and it is the secret author of many of the choices and judgements you make' (Kahneman 2011, 13).

The systematic improvement of assessment tasks requires the analytical approach of System 2, operating within a supportive environment where appropriate assessment-related decisions are made at the macro and meso levels. In such an environment, factors such as the socio-cultural context of assessment planning, institutional resourcing, reward structures,

departmental policies and disciplinary predilections are likely to be functioning to support, or at least not hinder, change. Even under such facilitative circumstances, however, the 'rationality' that is often associated with prescriptions for improving assessment can be limited by a variety of factors, including academics' time, their capacity to deal with only a limited number of ideas about improving assessment at any point, and their ability to translate advice into practice – factors encapsulated in the term, 'bounded rationality' (Simon 1955). These factors are nevertheless ones that can, under the right circumstances, be addressed by the conscious, deliberate efforts of System 2.

While recognising the broader cultural and organisational context of assessment change, this paper focuses on limits to change that can arise from factors associated with System 1 and which, because of their unconscious nature and the failure to date to relate these factors to decision making in assessment design, have received little attention in the higher education literature. These factors operate in a context where many, if not most, decisions are more conscious, deliberate and systematic, including decisions made at the department level as well as the more thoughtful decisions at the level of individual assessment tasks (Bearman et al. 2016; Boud et al. 2016). Consequently the factors which are the focus of this paper are only one set of factors potentially in operation within the broader decision-making context, but it is within this set that many influential decisions about whether and how to change assessment tasks may well lie.

Unconscious factors inhibiting change

Research over the past forty years into unconscious factors that can distort judgement and decision making have led to the 'heuristics and biases' understanding of decision making which now sits alongside analytical decision making as one of the two most prominent approaches to decision making in the literature (Fox 2014; Phillips, Klein, and Sieck 2014). The factors addressed in this section are drawn from that literature in which 'heuristics' refers to unconsciously and rapidly employed 'rules of thumb' to make judgements or decisions and which, while often serving us well, can also distort judgement, resulting in 'bias', that is, a somewhat predictable deviation towards one decision rather than another decision which a more deliberative, analytic process would have produced. Bias in this sense is not concerned with prejudice or predispositions but is merely an outcome of a misleading cognitive process.

While dozens of heuristics and biases are discussed in the literature, a smaller number have been subject to intense scrutiny (Kahneman, Slovic, and Tversky 1982; Gilovich, Griffin, and Kahneman 2002; Koehler and Harvey 2004). Our concern is with those that might compromise the process of improving assessment tasks in one of three ways: by supporting a reluctance to change existing tasks, by limiting the time allocated to develop alternative assessment tasks, or by underestimating the degree of change needed. Five such factors will be considered in this section: 'status quo bias' - the unjustified tendency to avoid change by retaining the status quo; the 'planning fallacy' heuristic which entails underestimating the time a task will take, leading to decisions being compromised by a lack of time and/or information; the 'anchoring and adjustment heuristic' according to which decisions can be unduly influenced by an initial starting point and failure to move as far from that as the situation warrants; 'the endowment effect' and the 'IKEA effect' involving overvaluing tasks due to a sense of ownership or an act of creation or construction; and 'overconfidence' when we overvalue our decisions based on our sense of having a coherent explanation for them regardless of how well founded the decisions may be. While none of these factors have been empirically researched in the context of planning assessment tasks, given how strongly they have been demonstrated elsewhere, it is appropriate to suggest how they might influence efforts to improve assessment tasks, how potential deleterious influences might be mitigated, and what research should be undertaken to establish their relevance or otherwise.

Status quo bias

> *Jim teaches a large first-year management course. He has just returned from a workshop on assessment where he learnt about several new assessment types. Now it's time for Jim to write his assessment for next semester. Some of these new tasks sound more reliable, more valid, and better for learning than his usual essay – and they might involve less effort to mark. But then again, first-year management has always had an essay, and it works well enough.*

The tendency not to change assessment tasks is most vividly exemplified by 'status quo bias' which inclines us, when faced with a number of choices, to select an option which has been framed as the status quo rather than opting for the best possible option. This occurs in contexts where there appears to be no noticeable advantage in this choice so that the framing of an option as the status quo is the only reason for its being preferred (Samuelson and Zeckhauser 1988). The status quo bias has been demonstrated experimentally through both simple and complex scenarios as well as in real life choices with respect to university employees' decisions regarding health plans and retirement funds. The results of experiments have shown the response to a particular option is highest when it is designated as the status quo, lower when it is in a neutral position (with no option designated as the status quo), and lowest when it is an alternative to the nominated status quo (Samuelson and Zeckhauser 1988). Numerous studies since 1988 have confirmed status quo bias in a wide variety of contexts and its practical application in commerce is rife, including invitations to change mobile phone plans or to accept special magazine subscription offers which after a given period turn into more expensive 'status quos' from which one must opt out (Thaler and Sustein 2009).

Academics in contexts where assessment change is viable may be aware of a small number of optional assessment methods; in most such situations, one method will be clearly designated as the status quo, namely the method currently being used. Status quo bias suggests that this method will have an unconscious attraction that is unrelated to its worth. It is interesting to note that in the United States, 'alternative assessment' is often positioned in contrast to mainstream assessment that is taken to primarily involve multiple-choice tests, thus positioning almost all assessment practices in contrast to the status quo (see, for example, Gipps and Stobart 2012).

A further issue arises when we are confronted by a large number of options. Efforts to improve assessment often expose academics to a wide range of options. However, it appears that as the number of options increase, so too does status quo bias, so that 'whereas subjects may be able to discriminate clearly between two alternatives according to their true preferences, discrimination among four would be more difficult, making the status quo more attractive as the path of least resistance' (Samuelson and Zeckhauser 1988, 25). Paradoxically, according to this conclusion, the more assessment options available, the less likely any of them will be adopted.

Of course, there may be good, 'rational' reasons for maintaining an existing set of assessment tasks. For example, the cost of making a change may seem to outweigh the perceived benefits, or navigating the assessment change system in a department may be judged to be too time consuming or troublesome. Where a course and its assessment have been 'inherited' from another teacher, it might be presumed that the existing assessment was well thought through previously and that further work on it is not warranted. Finally, there may be too much uncertainty about whether the benefits of alternative assessment tasks will accrue in practice - improving assessment often involves adopting methods that a teacher has not tried before, which therefore entail some risk, and which may lead to regret if the new tasks are not sufficiently effective (Kahneman and Tversky 1982a ; Kahneman 2011, 346-349). The combination of factors supporting the status quo leads to a rather discouraging conclusion:

> …the range of explanations for the existence of status quo bias …
> suggests that this phenomenon will be far more pervasive in actual
> decision making than the experimental results alone would suggest. The
> status quo bias is best viewed as a deeply rooted decision-making practice
> stemming partly from a mental illusion and partly from psychological
> inclination'. (Samuelson and Zeckhauser 1988, 10)

Nevertheless there may be steps that can be taken to reduce the influence of this bias when promoting assessment change. A starting point may be to simply increase academics' awareness of alternatives – in the absence of alternatives, the status quo may be considered the natural state of things. This could be combined with a requirement to justify 'no change' to existing assessment tasks, whereas institutional policies often reinforces the status quo by requiring justification only when change is contemplated. Altering the status quo 'status' of current assessment regimes by re-framing them as one of many possibilities should also be considered, keeping in mind the need to limit the number of options under consideration lest the status quo bias be reinforced. Finally, it would seem important to reduce the fear of regret by rewarding innovation and modifying student evaluation of teaching to avoid criticism or punishment if a new assessment task is not effective or popular.

The planning fallacy

> *Jane teaches a second-year history course. Several months ago a ruling came down from on high that all courses must use rubrics, and they must be distributed at the start of the semester. Jane didn't want to do one of those box-ticking analytic rubrics, and instead wanted something more holistic. She also knew that making a good rubric required feedback from students and her peers on the rubric, and piloting it with markers. An experienced colleague advised her this would take a couple of months, but to Jane it sounded more like two weeks' worth of work. Now it's two weeks before the start of semester. Jane is rapidly discovering she needs to learn more about rubrics before she can make one. A week later, with the completed rubric in hand, nobody is available to give her feedback, and her markers say they don't have time to pilot. At the end of the week, Jane uploads her box-ticking rubric to her course website, which is yet to be seen by anybody else.*

Significant changes to assessment tasks invariably require a degree of planning. A teaching team which completes its curriculum documentation by working late into the night the day before it is due has likely fallen victim to the planning fallacy, defined as the tendency 'to underestimate the time required to complete a project, even when we have considerable experience of past failures to live up to planned schedules' (Kahneman and Tversky, 1982b). The planning fallacy can be observed in major public building and infrastructure projects as well as in more prosaic instances such as academics writing a text book (Kahneman 2011, 245-247) or psychology students completing their masters dissertations (Buehler, Griffin, and Peetz 2010. The planning fallacy is not merely underestimating how long a task will take, but doing this when we have reliable information from our own or others' experience that the task will take much longer. Optimistic estimates of how long we think a task will take are said to be based on an 'internal' view, while taking into account others' or even our own experience in other similar situations is utilising an 'external' view (Kahneman and Tversky 1982b).

The planning fallacy would seem to have one major implication for assessment design: if significant improvements to multiple aspects of assessment are desired, the time required to plan these improvements is likely to be underestimated, leading to rushed work as a deadline approaches. If assessment revisions are part of a wider curriculum review of a course and if assessment is left to the end of this revision, the time available to work on assessment will be even further reduced. As available time decreases, the opportunities to seek out information,

for example alternative assessment methods, or to re-design specific practices such as the role and use of feedback, may be severely compromised.

Research on the planning fallacy has led to numerous suggestions for its avoidance (Kahneman and Tversky 1982b; Buehler, Griffin, and Peetz 2010). A critical step for academics or teaching teams redesigning assessment tasks would be to estimate time needed based on past experience of similar tasks, not on an overly optimistic forecast of their forthcoming work, that is, they should adopt an 'outside' view. The outside view can be strengthened by imagining how the task will unfold as if one were an outside observer. Another recommended process is to break the design/re-design of assessment into sub-tasks, thereby providing a reminder of the task's complexity and thus of the time needed for it. Perhaps the strongest way of addressing this problem is to consider assessment tasks sooner rather than later in the design or re-design process.

Of course, in many instances administrative systems and/or excessive workloads, can lead to courses being allocated to academics too late for meaningful change – the planning fallacy can be compounded by an actual shortage of time.

Anchoring and adjustment

> *Peter teaches third-year physics and has just returned from a physics education conference. He is now on a mission to completely redevelop the assessment in his course, with an emphasis on in-class assessment rather than exams. He glances up on his wall to last semester's syllabus document, which mentions an 80% exam. Several hours later he is proud of his new assessment design, which only weights the exam at 70%.*

It was noted in relation to status quo bias that a common starting point for improving assessment tasks is the set of tasks currently in place. The anchoring and adjustment heuristic draws further attention to the powerful influence of such starting points, the need to move away from them in order to reach a better decision about tasks, and the frequent failure to move far enough to achieve our goals. In its initial formulation, this heuristic was explained as follows:

> In many situations, people make estimates by starting from an initial value that is adjusted to yield the final answer. The initial value, or starting point, may be suggested by the formulation of the problem, or it may be the result of a partial computation. In either case, adjustments are typically insufficient … That is, different starting points yield different estimates, which are biased toward the initial values. We call this phenomenon anchoring. (Tversky and Kahneman 1974, 1128)

Most research on anchoring and adjustment has been numerically based. For example, in re-marking student essays, a knowledge of the first mark has been found to act as an anchor for the second marker, with the resulting grade being closer to the first than if this had not been known (Brookes 2012). This heuristic seems to apply to a wide range of decisions, to the extent that it has been described as 'ubiquitous in human judgment' (Eply 2004, 253) while Furnham and Boo, in their review of anchoring effect literature, conclude that '… the influence of anchoring is exceptionally robust, pervasive and ubiquitous' (Furnham and Boo 2010, 41) and that it 'has many implications in all decision making processes' (Furnham and Boo 2010, 35).

The potential implications of this heuristic for assessment change are clear: if current assessment tasks act as an anchor, an academic who may otherwise design quite different tasks may adjust what they consider most appropriate towards what is currently in place.

Resisting the anchoring and adjustment effect may not be easy, given its robustness and its unconscious nature; even if we are aware of the presence of anchors, we are not aware of the

influence they have on our judgement or decisions. Furnham and Boo, in an economic context, suggest broadening the pool of advisors, doing more thorough research, and looking at trends (Furnham and Boo 2010, 41). Assessment design equivalents might be to seek external advice from colleagues with different approaches to assessment and to consult literature on a wide range of assessment options within and outside one's discipline.

The Endowment Effect and the IKEA Effect

> *Jenny teaches first-year physiology. She made significant changes to one assignment last year, which required substantial time and effort. Jenny liked the task she produced, though it seemed unpopular with students and colleagues; it's also a lot of work to mark. An email just came through on a physiology learning and teaching list she subscribes to, about a new type of task which promises to be much easier to mark, and wildly popular with everybody involved. It also seems to have a good evidence base about its reliability and validity. Jane could feasibly just copy the task over to her unit and make a few small changes. But she likes her task better.*

Teachers in higher education often display a strong sense of ownership regarding their work, including their ways of assessing their students' learning. The endowment effect refers to such a propensity to value something owing to a sense of ownership of it, regardless of its inherent value or how we came by it (Kahneman, Knetsch, and Thaler 1990; Thaler 2015). The endowment effect is closely associated with 'loss aversion' – once we 'own' something our aversion to losing it is greater than the pleasure associated with gaining something equivalent or even better (Kahneman, Knetsch, and Thaler 1990; Kahneman 2011).

How might this tendency contribute to a reluctance to change assessment? Ownership of assessment tasks might arise because of the academic's role in designing the course, including its assessment (in which case it may be subject to the IKEA effect discussed below), but it may also arise when a course and its assessment tasks are 'inherited' from a colleague or where the tasks are simply consistent with departmental or disciplinary norms with which the academic identifies. While the inherited tasks may have seemed unexceptional previously, once they are part of one's responsibility, they may well be valued more highly. Moreover, they may be valued more highly than other possible assessment tasks that would, in other respects, be considered equally or even more appropriate (Kahneman, Knetsch, and Thaler 1991; Thaler 2015).

The IKEA effect refers to a similar over-valuation of something for which we have a sense of ownership, but now the over-valuation is associated with one's having built or put it together (Norton, Mochon, and Ariely 2011). While assembling IKEA furniture figured in an initial experiment, the implications of the IKEA effect are thought by its instigators to be much broader. For example, the IKEA effect has been related to the 'not invented here' syndrome whereby 'managers refuse to use perfectly good ideas developed elsewhere in favor of their – sometimes inferior – internally-developed ideas' (Norton, Mochon, and Ariely 2011, 21).

If the IKEA effect applies to assessment, academics may over-value assessment regimes they have developed themselves and may see limited need to change them. Moreover, this tendency may be higher the more innovative the assessment regime is. Such assessment is likely to be closely aligned with the teacher's conceptions of learning and teaching and of the discipline itself, making it even more resistant to change. While unconscious effects are difficult to counter by their very nature, theorists have posited ways of compensating for many of them. While this has not occurred so far with the IKEA effect, one might postulate that a more distributed responsibility for the design of assessment tasks which does not locate them as the possession of an individual academic may ameliorate some aspects of this effect.

Overconfidence

> *Barry teaches second-year electrical engineering. He is quite happy with his assessment: every semester it produces a very nice grade distribution which satisfies his department without the need to scale student marks. Although Barry's colleagues are changing their tasks based on some other assessment principles he doesn't fully understand, Barry is confident he still has high-quality assessment. The grades look fine – what more could he want?*

The confidence a teacher or teaching team has in their assessment tasks would be expected to directly reflect the quality of those tasks in meeting their particular purposes. Yet within heuristics and biases research, a sense of confidence is seen to be potentially quite misleading:

> Subjective confidence in a judgment is not a reasoned evaluation of the probability that this judgment is correct. Confidence is a feeling, which reflects the coherence of the information and the cognitive ease of processing it. It is wise to take admissions of uncertainty seriously, but declarations of high confidence mainly tell you that an individual has constructed a coherent story in his mind, not necessarily that the story is true. (Kahneman 2011, 212)

Rather worryingly, Kahneman notes that '(p)aradoxically, it is easier to construct a coherent story when you know little, when there are fewer pieces to fit into the puzzle. Our comforting conviction that the world makes sense rests on a secure foundation: our almost unlimited ability to ignore our ignorance' (Kahneman 2011, 201). In the presence of the overconfidence effect, confidence in an assessment regime may not reflect its quality – indeed, the opposite may be the case, and an inclination to not change that regime may be stronger the less well informed about assessment an academic is. Confidence in an existing assessment regime would be a powerful reason to not change it, and if that confidence rests on a mistaken belief of its adequacy arising from limited information about assessment, some of the most poorly designed assessments may be the most impervious to change.

Three ways to address overconfidence in assessment planning can be found in the decision making literature. 'Counterargumentation' involves thinking of reasons why planned assessment tasks may not be the most appropriate, thus prompting a re-consideration. 'Paths to trouble' or 'fault trees' mapping the points in the assessment implementation cycle where problems might arise (Russo and Schoemaker 1992), a process sometimes referred to as 'pre-mortems' (Soll, Milkman and Payne 2015), can similarly prompt further consideration.

Countering unconscious limits to change

If the above five factors do indeed influence decisions about changing assessment tasks, what can be done to mitigate their influence? In the previous section a number of specific ways of countering each of these factors were noted. None of these suggestions provide ways of changing a teacher's way of thinking but rather depend on balancing limitations by using additional, conscious strategies. An important reason for this is the apparently unchangeable nature of the unconscious processes. The conclusion that 'System 1 is not readily educable' (Kahneman 2011, 416) and that 'there are many reasons to doubt that lone individuals can debias themselves' (Larrick 2004, 318) seems well supported by decision-making research (Larrick 2004). The best that can be hoped for is that one can come to recognise situations where decision making might be distorted and deliberately apply more careful, analytical processes, though the capacity to actually do this seems limited since this depends on (a) the undoubtedly limited capacity to recognise such situations and (b) the teacher having a repertoire of strategies for identifying and considering his or her assessment options. The

latter can be supported by access to information, including exemplars and frameworks for assessment decisions (Bearman et al. 2016). Such tools and resources to develop assessment literacy are commonly available in higher education institutions, either through academic development units or online guides to various aspects of assessment tasks, and of course there is an extensive, though not frequently accessed, literature on assessment.

A more promising approach to decision making in the context of improving assessment tasks lies in what has been termed 'decision architecture' or 'choice architecture' (Thaler and Sustein 2009; Beshears and Gino 2015; Soll, Milkman and Payne 2015), that is, creating an environment that supports or requires informed and thoughtful approaches to decisions. For example, providing incentives for improving assessment tasks may reduce the influence of anchors (Larrick 2004, 322). Requiring accountability, whereby decisions about assessment tasks must be explained to others, may lead to teachers giving assessment planning greater attention, anticipating limitations to their plans, and seeking and using more information. Within an accountability framework, 'considering the opposite', that is, forms of assessment quite different to the status quo, or proposing other alternatives, all promote more careful thought of a wider range of options (Larrick 2004: Soll, Milkman and Payne 2015). While peer review of assessment is usually conceived of as a cross-institutional moderation of student work (Bloxham, Hudson, den Outer and Price 2015), informal reviews of assessment tasks within a department or discipline, combined with reviews by departmental assessment committees, might be even more effective in improving the quality of assessment tasks.

An important key to addressing limits to improving assessment would be to require justification of decisions to leave assessment unchanged. In many higher education institutions, an elaborate justification process is required to change assessment, but usually no justification is required to leave an assessment regime as it is – an outstanding example of poor decision architecture which could be easily remedied.

Of course, requirements to consider alternatives, to engage in peer or departmental review of assessment tasks, and to justify assessment regimes are all time-consuming. Consequently they could easily lead to resentment or superficial compliance and are likely to be successful only in contexts where assessment is seen as central to teaching and learning and the improvement of assessment tasks is consequently highly valued and supported.

Researching assessment task change

While the unconscious decision making processes discussed in this paper have been widely researched and documented in the specialist field of judgement and decision making and in relation to a range of contexts, they have received little attention in educational contexts and their application to assessment task design, as far as we are aware, has not previously been researched. With assessment being such a critical issue in higher education, and with imperatives to change assessment coming from many stakeholders, there is a pressing need to better understand how such change can be brought about, what factors might impede it, and how these might be overcome. While the issues raised in this paper would seem to have important implications for improving assessment design, little research is available to indicate if or how they apply to assessment decisions. Consequently the scope for research is considerable.

The underlying question relating to each of the heuristics and biases is simply 'Does this arise when academics are considering changes to their assessment tasks?', a question which sits in the broader context of how academics experience the process of changing, or not changing such tasks. For each factor, a number of more specific questions arise. For example, regarding status quo bias, we might want to consider the frequency of 'no change' decisions in assessment tasks, the reasons for such decisions, how academics 'inheriting' a course and its assessment perceive the existing assessment regime and the influence this has on their

willingness or otherwise to change the assessment. In considering the planning fallacy, we might want to compare how much time an academic or teaching team expect to spend on revising assessment with actual time spent and the effects of any discrepancies. An extensive list of such questions could readily be developed in relation to each of the factors addressed in this paper.

Conclusion

The question 'Why is it so hard to change assessment?' leads in many directions. The issues canvassed in this paper suggest a partial explanation of a complex phenomenon – certain unconscious factors in the decision-making of individuals limit openness to change, reduce the scope of change, or compromise the capacity to plan for change in a timely way.

Future interventions to improve assessment at the course level would be well advised to pay attention to the heuristics and biases noted in this paper, while the possible implications of others remain to be considered. This paper has drawn attention to just some of those factors which have the potential to exert an influence on assessment planning and design by inhibiting or limiting essential improvement. Whether these factors actually do so needs to be established. An understanding of how they function as barriers to change is an essential first step in reducing their influence. While this paper's argument is for better decision making processes, the actual decisions that result will only serve to improve the experience of students and academics if they lead to assessment tasks designed to meet the multiple functions of assessment outlined at the start of the paper.

Acknowledgements

References

Bearman, M., P. Dawson, D. Boud, S. Bennett, M. Hall, and E. Molloy. 2016. "Support for Assessment Practice: Developing the Assessment Design Decisions Framework." *Teaching in Higher Education*, 21:5, 545-556.

Beshears, J., and F. Gino. 2015. "Leaders as Decision Architects." *Harvard Business Review* May: 52 – 62.

Bloxham, S., J. Hudson, B. den Outer, and M. Price. 2015. "External Peer Review of Assessment: An Effective Approach to Verifying Standards?" *Higher Education Research & Development*, 34:6, 1069-1082. doi: 10.1080/07294360.2015.1024629

Boud, D. and Associates. 2010. *Assessment 2020: Seven Propositions for Assessment Reform in Higher Education.* Sydney: Australian Learning and Teaching Council. Accessed 1 April 2016. http://www.uts.edu.au/sites/default/files/Assessment-2020_propositions_final.pdf .

Boud, D., P. Dawson, M. Bearman, S. Bennett, G. Joughin, and E. Molloy. 2016. "Reframing Assessment Work: through a Practice Perspective." *Studies in Higher Education*. doi.org/10.1080/03075079.2016.1202913

Boud, D., and N. Falchikov. 2007. "Developing Assessment for Informing Judgement." In *Rethinking Assessment in Higher Education*, edited by D. Boud and N. Falchikov, 181-197. London: Routledge.

Bromage, A. 2006. "The Management of Planned Change: An Interdisciplinary Perspective." In *The Realities of Change in Higher Education: Interventions to Promote Learning and Teaching*, edited by L. Hunt, A. Bromage, and B. Tomkinson, 3-13. New York: Routledge.

Brookes, V. 2012. "Marking as Judgment." *Research Papers in Education* 27 (1): 63-80.

Buehler, R., D. Griffin, and J. Peetz. 2010. "The Planning Fallacy: Cognitive, Motivational, and Social Origins." *Advances in Experimental Social Psychology* 43: 1-62.

Commonwealth of Australia. 2015. *Higher Education Standards Framework (Threshold Standards) 2015*. Accessed 9 June 2016. https://www.legislation.gov.au/Details/F2015L01639.

Deneen, C., and D. Boud. 2014. "Patterns of Resistance in Managing Assessment Change." *Assessment & Evaluation in Higher Education* 39 (5): 577-591.

Eply, N. 2004. "A Tale of Tuned Decks? Anchoring as Accessibility and Anchoring as Adjustment." In *Blackwell Handbook of Judgment and Decision Making*, edited by D. J. Koehler, and N. Harvey, 240-257. London: Blackwell.

Ewell, P. T. 2013. *The Lumina Degree Qualifications Profile (DQP): Implications for Assessment.* Champaign: National Institute for Learning Outcomes Assessment. Accessed 1 April 2016. http://www.learningoutcomesassessment.org/documents/EwellDQPop1.pdf.

Fox, J. 2014. "From 'Economic Man' to Behavioural Economics." *Harvard Business Review* 93 (5): 78-85.

Furnham, A., and H. C. Boo. 2010. "A Literature Review of the Anchoring Effect." *The Journal of Socio-Economics* 40: 35–42.

Gilovich, T., D. Griffin, and D. Kahneman, eds. 2002. *Heuristics and biases: the psychology of intuitive judgment.* Cambridge: Cambridge University Press.

Gipps, C., and G. Stobart. 2012. "Alternative Assessment." In *International Handbook of Educational Evaluation*, edited by T. Callaghan, D. L. Stufflebeam, and L. A. Wingate, 549-575. Dordrecht: Springer.

Joughin, G. ed. 2009. *Assessment, Learning and Judgement in Higher Education*. Dordrecht: Springer.

Kahneman, D. 2011. *Thinking, Fast and Slow*. London: Penguin.

Kahneman, D., J. L. Knetsch, and R. Thaler. 1990. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy* 98 (6):1325–1348.

Kahneman, D., J. L. Knetsch, and R. H. Thaler. 1991. "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias." *The Journal of Economic Perspectives* 5 (1): 193 – 206.

Kahneman, D., and A. Tversky. 1982a. "The Psychology of Preferences." *Scientific American* 246 (1): 160-173.

Kahneman, D., and A. Tversky. 1982b. "Intuitive Predication: Biases and Corrective Procedures." In *Judgment under Uncertainty: Heuristics and Biases*, edited by D. Kahneman, P. Slovic, and A. Tversky, 415-421. Cambridge: Cambridge University Press.

Kahneman, D., P. Slovic, and A. Tversky, eds. 1982. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Koehler, D. J., and N. Harvey, eds. 2004. *Blackwell Handbook of Judgment and Decision Making*. Blackwell Publishing Ltd, Malden, MA, USA

Kreber, C., C. Anderson, N. Entwistle, and J. McArthur, eds 2014. *Advances and Innovations in University Assessment and Feedback*. Edinburgh: Edinburgh University Press.

Larrick, R. P. 2004. "Debiasing". In *Blackwell Handbook of Judgment and Decision Making*, edited by D. J. Koehler, and N. Harvey, 316-337. London: Blackwell.

Macdonald, R., and G. Joughin. 2009. "Changing Assessment in Higher Education: A Model in Support of Institution-wide Improvement." In *Assessment, Learning and Judgement in Higher Education*, edited by G. Joughin, 193–214.

Merry, S., M. Price, D. Carless, and M. Taras. 2013. *Reconceptualising Feedback in Higher Education: Developing Dialogue with Students.* London: Routledge.

Norton, M. I., D. Mochon, and D. Ariely. 2011. *The 'IKEA Effect': When Labor Leads to Love*. Harvard Business School Working Paper 11-091. Accessed 1 April 2016. http://www.hbs.edu/faculty/Publication%20Files/11-091.pdf.

Phillips, J. K., G. Klein, and W. R. Sieck. 2004. "Expertise in Judgment and Decision Making: A Case for Training Intuitive Decision Skills." In *Blackwell Handbook of Judgment and Decision Making*, edited by D. J. Koehler, and N. Harvey, 297-315. London: Blackwell.

Price, M., B., C. O'Donovan, C. Rust, and J. Carroll. 2008. "Assessment Standards: A Manifesto for Change." *Brookes eJournal of Learning and Teaching* 2 (3). Accessed 20 April 2016. http://bejlt.brookes.ac.uk/article/assessment_standards_a_manifesto_for_change

Russo, J. E., and P. J. Schoemaker. 1992. "Managing Overconfidence." *MIT Sloan Management Review*, Winter, January.

Samuelson, W., and R. Zeckhauser. 1988. "Status Quo Bias in Decision Making." *Journal of Risk and Uncertaint*y 1 (1): 7-59.

Simon, H. 1955. "A Behavioural Model of Rational Choice." *Quarterly Journal of Economics* 69: 99-11.

Soll, J. B., K. L. Milkman, K. L., and J. W. Payne. 2015. "Outsmart Your Own Biases." *Harvard Business Review*, May: 65-72.

Stanovich, K. E., and R. F. West. 2000. "Individual Differences in Reasoning: Implications for the Rationality Debate?" *Behavioral and Brain Sciences* 23: 645-726

Thaler, R. H. 2015. *Misbehaving: The Making of Behavioral Economics*. London: Allen Lane.

Thaler, R. H., and C. R. Sustein. 2009. *Nudge*. London: Penguin.

The Quality Assurance Agency for Higher Education. 2014. *The UK Quality Code for Higher Education Subject benchmark statements.* Accessed 30 May from http://www.qaa.ac.uk/assuring-standards-and-quality/the-quality-code/subject-benchmark-statements

Trowler, P., L. Fanghanel, and T. Wareham, 2005. "Freeing the Chi of Change: the Higher Education Academy and Enhancing Teaching and Learning in Higher Education." *Studies in Higher Education* 30 (4): 427-444.

Trowler, P., M. Saunders, and P. Knight. 2003. *Change Thinking, Change Practices*. York: Higher Education Academy. Accessed 9 May 2016. https://www.heacademy.ac.uk/sites/.../id262_change_thinking_change_practices.pdf.

Tversky, A. and D. Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185: 1124-1131.