

The mechanism of impact of summative assessment on medical students' learning

Francois J. Cilliers · Lambert W. Schuwirth · Hanelie J. Adendorff ·
Nicoline Herman · Cees P. van der Vleuten

Received: 23 November 2009 / Accepted: 26 April 2010 / Published online: 9 May 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract It has become axiomatic that assessment impacts powerfully on student learning, but there is a surprising dearth of research on how. This study explored the mechanism of impact of summative assessment on the process of learning of theory in higher education. Individual, in-depth interviews were conducted with medical students and analyzed qualitatively. The impact of assessment on learning was mediated through various determinants of action. Respondents' learning behaviour was influenced by: appraising the impact of assessment; appraising their learning response; their perceptions of agency; and contextual factors. This study adds to scant extant evidence and proposes a mechanism to explain this impact. It should help enhance the use of assessment as a tool to augment learning.

Keywords Assessment · Learning · Mechanism of impact · Determinants of action

Introduction

I think, as a result of assessment, I know much less than I otherwise would have.
(6(V)F65)

It has become axiomatic to refer to the powerful impact of assessment on student learning. Terms like the “backwash effect” (Biggs 1996; Elton 1987, used in general educational literature) and the “washback effect” (Alderson and Wall 1993; Bailey 1996, used in language teaching and testing literature) of assessment, “consequential validity” (Boud 1995), “test-enhanced learning”, the “testing effect” or the “testing phenomenon” (Glover 1989; Roediger and Karpicke 2006) and “test expectancy” (Lundeberg and Fox

F. J. Cilliers (✉) · H. J. Adendorff · N. Herman
Stellenbosch University, Stellenbosch, Western Cape, South Africa
e-mail: fjc@sun.ac.za

L. W. Schuwirth · C. P. van der Vleuten
University of Maastricht, Maastricht, The Netherlands

1991) have been used in this regard. A heritage of literature stretching back almost a century (e.g., Jones 1923; Meyer 1936) is widely cited in support of this phenomenon.

The impact of assessment on student learning is generally held to be profound. Elton and Laurillard (1979) went so far as to state that “the quickest way to change student learning is to change the assessment system”. Boud et al. (1999) state that “[a]ssessment is the single most powerful influence on learning in formal courses”. If this is the case, then assessment may well be one of the most powerful tools we have at our disposal to influence student learning. However, even after a almost a century of research, efforts to positively influence learning through assessment do not always yield encouraging results (Gijbels et al. 2009). In fact, how little we actually know about the complex relationship between assessment and student learning has been pointed out from various quarters (Alderson and Banerjee 2001; Lundeberg and Fox 1991; Ramsden 2005; Segers and Dochy 2006).

Different authors write about different things when they write about the impact of assessment on student learning. As far as “assessment” goes, authors variously focus on approaches like formative assessment, continuous assessment or coursework, internal-to-programme summative assessment and external-to-programme standardised testing, or on specific methods. Internal-to-programme summative assessment may well exert both a stronger and a more pervasive influence on learning than other assessment practices in higher education (HE), however. As Boud (1995) highlighted, students cannot escape the impact of summative assessment. Given the stakes, the design of such assessment is more typically informed by psychometric than learning considerations and so even if other aspects of assessment in a course have been designed to promote meaningful learning, the impact of summative assessment could trump beneficial effects achieved by other means. Furthermore, more students in HE probably encounter internal-to-programme summative assessment than external-to-programme standardised testing.

As for “learning”, it is often not the only phenomenon that authors highlight when writing about the impact of assessment. Models of language testing washback (Alderson and Banerjee 2001; Bailey 1996) and measurement driven instruction (Airasian 1988; Madaus 1988) both address the multifactorial impact of standardised testing on learning, teaching, materials, curricula and even research, though focussed largely on school-based settings. In HE, the impact of assessment on not just learning but also on non-learning student behaviours intended to enhance marks e.g., ingratiating themselves with lecturers, on student stress and on students’ choice of courses has been highlighted by various authors (Becker et al. 1968; Miller and Parlett 1974; Snyder 1971). From the opposite perspective, assessment is typically identified as one of the contextual factors that impact on learning in models of learning (Biggs 1987; Ramsden 1984; Ross et al. 2003; Vermunt 1996). Even when writing specifically about the impact of assessment on learning, authors variously write about the relationship between assessment and the product or outcome of student learning i.e., student performance and about the impact on the process of student learning.

Demonstrating a positive impact of assessment on desired student performance i.e., the outcome of student learning, may well be the ultimate goal of utilising assessment to enhance learning. However, efforts to effectively use assessment to achieve this would be enriched by an understanding of the intervening process(es) by which assessment impacts on the process of student learning. There is some descriptive literature focussing on the “what” of the impact of internal-to-programme summative assessment on the process of student learning in HE i.e., the sources and consequences of the impact of such assessment or, as Maxwell (2004a: 4) puts it, “*whether* x caused y” (emphasis in original). In contrast, very little has been written explaining “*how* it did so” (Maxwell 2004a: 4, emphasis in

original). As Bunge (2004: 199) noted, “any mechanism-free account must be taken to be shallow and therefore a challenge to uncover unknown mechanism(s)”. At issue in the present study is how summative assessment brings about its influence on the process of student learning. From a process theory perspective, this paper deals with “events and the processes that connect them” (Maxwell 2004b: 248) specifically in one distinctive, internal-to-programme, high-stakes assessment system; thus, the local mechanisms at play in a “complex network of events and processes in a situation” (Miles and Huberman 1994: 146; 147).

So what is known about the mechanism(s), the “how”, of impact? Such literature as there is has, for the most part, accumulated piecemeal, often coincidentally during work addressing broader issues. Given the difficulties generalising findings from controlled settings to the classroom (Lundeberg and Fox 1991), research conducted in controlled settings will not be reviewed here.

Extrinsic motivation

Assessment provides extrinsic motivation and impacts on the amount and distribution of students’ learning efforts. The mere fact of assessment motivates students to learn and therefore influences the quantum of effort expended on learning (Miller and Parlett 1974; Snyder 1971; van Etten et al. 1997). The impact of assessment on effort is not necessarily always positive, however. If students perceive they are unable to successfully negotiate assessment, for example because their marks are so bad they cannot hope to achieve a pass or they are so far behind they believe they cannot catch up, this can result in them stopping learning (Becker et al. 1968). Nor is the response to assessment as extrinsic motivation uniform. Sambell and McDowell (1998) described two case studies illustrating how differences in student motivation and student conceptions of learning elicit differing responses in different individual students to the same assessment task.

Consequences

The potential consequences of assessment also impact student learning. Students adapt both what and how they learn so as to meet the lecturers’ requirements as manifested in assessment rather than understand the material being learned (Becker et al. 1968; Ramsden 1984, 1992; Snyder 1971). This is at least in part because the risks of not doing so are great, the rewards for conforming, substantial, both in terms of self-esteem and in terms of short and longer term material benefits (Parlett 1969; Snyder 1971). The likelihood of subject matter featuring in assessment impacts on what content students select to learn (Becker et al. 1968; Becker et al. 1961; Miller and Parlett 1974; Snyder 1971; Vermunt 1996) or not—what Snyder (1971) referred to as “selective negligence”. It also influences the thoroughness with which students engage with learning material (Laurillard 1979; van Etten et al. 1997). This also holds for assignments students choose to do or not (Snyder 1971) and the amount of effort students devote to tasks (Becker et al. 1968; Janssens et al. 2002, cited by Struyven et al. 2005). The thoroughness with which students engage with learning material is impacted by the contribution that performance on any given assignment will make towards the calculation of a final grade (Ramsden 1992; Snyder 1971). Personal consequences like the risk of appearing ignorant in an oral assessment also impacted how thoroughly students prepare (Joughin 2007).

Achieving a desired outcome

The likelihood of any given learning behaviour bringing about a desired assessment outcome influences students' actions. The amount of time students spend studying increases, up to a point, as the volume of material and, independent of that, the degree of difficulty of the material, to be studied, increases (van Etten et al. 1997). Students select resources and activities to prepare for assessment that best prepare them for the demands of the assessment task (Frederiksen 1984; Newble and Jaeger 1983). Student also match the nature of their learning to the demands of the assessment task to achieve a desired outcome (Becker et al. 1968; Sambell and McDowell 1998). Scouller (1998) also offered some tantalizing evidence that perceptions of the cognitive demands of an assessment task are correlated with the approach to learning adopted, but could not comment on causality. Students preparing for an MCQ assessment task were both more likely to adopt a surface approach and perceive the assessment to be pitched at a lower cognitive level than when they were preparing for an essay assignment.

Students seek cues from lecturers, other students and past papers to guide their selection of content to learn, in the interests of achieving their desired outcome with assessment (Becker et al. 1968, 1961), and may even cheat to achieve this end (Becker et al. 1968). High volumes of work drive students to be more selective about what content to engage with and to adopt low level cognitive processing tactics in the interests of achieving a desired outcome (Ramsden 1984; Snyder 1971; van Etten et al. 1997). Effort is allocated across courses based on where generating benefit or reward is deemed most useful at any given time (Becker et al. 1968).

Goals

Students' goals influence their response to assessment. Students gauge the magnitude of their efforts by what grade they aim to achieve (Becker et al. 1968; Miller and Parlett 1974; van Etten et al. 1997). Various factors influence the priority students accord reading assignments, including whether they need to improve in the subject, whether the material is interesting, whether the material is manageable (e.g., not impossible to understand) and whether the assignment is in their major area of study (van Etten et al. 1997). Interestingly, the type of learning students adopt to meet the perceived demands of assessment may be discordant with their long-term goals. Entwistle and Entwistle (1992) referred to "the way in which the examination distorted the efforts of the students to achieve personal understanding"—students' learning was guided by the type of questions in past papers and thus the perceived requirements of old examination questions. Various authors describe how students experience a tension between learning what and how they would like to and learning what and how they need to, to succeed in examinations (Becker et al. 1968, 1961; Entwistle and Entwistle 2005; Ramsden 1992; Snyder 1971; Tang 1994).

Norms

Individual responses to assessment can be modulated by a socially constructed and shared frame of reference within a peer group (Becker et al. 1968). The norms within a peer group can modulate when a student starts learning, resulting in them starting later than they would otherwise have chosen to (Thomson and Falchikov 1998).

Agency

Beliefs about agency appear to mediate students' response to assessment. Students' beliefs as to whether studying would influence their performance on assessments affects their motivation to learn (van Etten et al. 1997). When students start learning is influenced by their perception their ability to cope with a task of given magnitude and complexity, given the prevailing workload (Snyder 1971). Low-level processing may also be a response to work deemed too complex to understand, when the effort to understand outweighs the potential reward of doing so rather than just memorising (Becker et al. 1968).

Emotion

Lastly, emotion also mediates students' responses to assessment. In addition to interest, Fransson (1977) also reported students' approach to learning is impacted by the degree of threat and anxiety they experience, both factors associated with assessment. Worry about assessment has also been reported to influence allocation of effort to learning (Miller and Parlett 1974).

Two things about this literature are striking. The first is that, for the most part, these studies were not designed to systematically investigate the impact of assessment on learning. The findings reported above mostly represent fragments from the data reported in these studies; thus, at best, examples of how assessment can influence learning, rather than a systematic record of how this comes about. There are various studies that are often cited as providing evidence of the impact of assessment on learning. However, many of these were experimental work conducted in controlled settings with limited ecological validity (Lundeberg and Fox 1991). Furthermore, many of these studies were conducted in school settings, including elementary schools, further limiting their usefulness in HE settings.

The second thing that is striking about this literature is that only limited attempts have been made to explain the impact of assessment within a theoretical framework. In the washback literature, various hypotheses (Alderson and Wall 1993) and a theoretical model (Bailey 1996) of washback have been proposed, but the impact of assessment on learning within this model enjoys only limited empirical support. Alderson and Wall (1993) did posit that the impact of assessment may result from an impact on motivation and therefore behaviour. Both van Etten et al. (1997) and Ross et al. (2006, 2003) invoke models of self-regulation. van Etten et al. (1997) subsequently propose a framework based on empirical data comprising a number of propositions divided into four categories about students' beliefs about preparing for examinations. Ross et al.'s (2006, 2003) subsequent exploration of the relationships between instruction, assessment, learning strategies and academic performance, while grounded in theory, did not address the role of motivation. Furthermore, such evidence as they offer is generated in an experimental setting. Becker et al. (1968) proposed the "grade point average perspective" which "describes the situation in which students see themselves working, the rewards they should expect from their academic work, the appropriate actions to take in various circumstances, the criteria by which people should be judged". Although they offer extensive evidence to support the existence of this perspective which, like the concept of washback, extends beyond the impact of assessment on learning, it is not tied to any underlying theoretical framework. Broekkamp and van Hout-Wolters (2007) propose and provide theoretical support for a model describing students' strategy adaptation when preparing for tests. Much of the literature they draw on in support of their model is school-based.

If assessment is to be used as a tool to enhance the power of the learning environment (de Corte et al. 2003) and contribute to, rather than mitigate, the cultivation of productive learning (de Corte 2007) in HE, we need to understand not only what impact assessment has but how that impact is brought about. The purpose of this study was, therefore, to investigate the impact of summative assessment on student learning and to specifically explore the mechanisms by which assessment impacts on learning. The key research question here was: How do various dimensions of summative assessment of theory bring about such influence as they exert on various dimensions of learning in a HE setting? This research was qualitative and exploratory in nature, aiming to inductively develop a model to start explaining the impact of assessment on student learning.

Methods

Context

This study was conducted at the Faculty of Health Sciences of Stellenbosch University in South Africa. Medical students there follow a 6 year, modular program. During the third of four phases of the program i.e., semesters four to nine, transdisciplinary, system-based theory modules (e.g., Cardiovascular System or Respiratory System, typically offered in lecture halls) alternate with discipline-based clinical modules (where students acquire clinical skills in various disciplines e.g., Surgery or Paediatrics, typically in clinical settings). Alternating 4 week periods of study are allocated to theory and clinical modules. Most modules are 4 weeks long, though some are shorter or longer. The final three semesters of the program, the fourth phase, comprise solely clinical modules. Students receive study guides for each module detailing module level and often also session level outcomes and detailed information on assessment procedures and requirements.

During Phase 3, summative assessment takes place for each module and assessment stakes are high. In theory modules—the focus of this study—marks generated by assessment during the module are combined with marks generated by an end of module assessment to generate what is called the class mark. If students fail to achieve a class mark of at least 40%, they do not qualify for access to the end of year examination in that module and therefore fail the year. In the year-end examination, students who score <45% fail the module outright and therefore fail the year. Students who score between 45 and 50% qualify for a resit. All students who fail to achieve $\geq 50\%$ in the resit, also fail the module and the year. If students fail any module, they have to repeat that year of study. Students may only repeat a year of study once during the program, although they may repeat more than one module in that year. Generally, attrition is about 25% during the course of the program and highest during the first 2 years.

Subjects and ethics

Ethical approval was obtained for the study from an institutional research ethics board. Based on the fact that they had had several semesters' experience being assessed on both theory and clinical skills, all students in the fourth and fifth years of the program were invited to participate. Each class was addressed once about the study and one email reminder was sent to each student. No incentive was offered for participation in the study.

Thirty-two students volunteered for interviews. Interviews were scheduled at the convenience of students and took place during students' seventh (for fourth year students) or

Table 1 Distribution of respondents based on year of study, gender and academic performance

Year of study	Gender	Average mark		
		<70%	70–79%	≥80%
4	F	13(IV)F69		
	M	7(IV)M72 16(IV)M79		
5	F	6(V)F65	2(V)F77	4(V)F81
		12(V)F61(1)	11(V)F71(1)	8(V)F81
		15(V)F63(3)		9(V)F83
	17(V)F65		18(V)F86	
M	3(V)M67	5(V)M73 14(V)M78	1(V)M80 10(V)M82	

The following notation will be used to identify respondents: **12(V)F61(1)**, where the first figure indicates respondent number; the Roman numeral in parentheses their year of study; the letter their gender; the next figure their average mark across all 6 years of study (obtained some time after the interviews); and the last parentheses indicate the number of modules (if any) that the respondent failed across all years of study

ninth (for fifth year students) semester of study. Eighteen respondents were interviewed. The remainder were thanked but not interviewed, given that data saturation had been achieved (see below). During interviews, informed consent for study participation and later access to students' academic records was elicited using an information sheet and informed consent document.

Some characteristics of respondents are summarized in Table 1. Given that approximately one-fifth of the two classes from which they were drawn achieved an average mark $\geq 70\%$ (data not shown), it is evident that this group includes a higher proportion of more successful students than the classes from which they are drawn. However, the number of students who failed one or more modules (see footnote to Table 1) indicates these students' learning was not uniformly effective. Two-thirds of respondents were women, a slightly higher proportion than the 55% of the two classes concerned that were women.

Data collection and analysis

In-depth, unstructured interviews (Charmaz 2006; DiCicco-Bloom and Crabtree 2006; Kvale 1996) were conducted with individual students, each lasting approximately 90 min. In keeping with the inductive nature of the study, no formal interview schedule was used. Interviews were loosely constructed around exploring three issues: how respondents learned, what assessment they had experienced and how assessment had impacted on their learning. Open-ended questions were used and statements respondents made were probed to clarify meaning, obtain additional detail and ascertain what assumptions underlie them. For example, vague statements like "I learn differently for long questions and multiple choice questions" were probed for detail about what respondents did differently in the two situations and why they did so.

Although the interviews were conducted at one point in time, students' experience of different assessment methods and how they learned in varying contexts across all of their years of study were explored, compared and contrasted during interviews, though typically not chronologically. This revealed qualitative and quantitative differences and changes in respondents' learning across varying assessment contexts and time. Each interview was

allowed to develop its own direction within the broad three-topic framework, so as to allow in-depth exploration of each respondent's experiences and conceptions of the relationships being studied. Given that data collection proceeded in tandem with, and was later informed by, data analysis, as analysis proceeded, emerging constructs were also discussed with respondents to confirm interpretation and explored in greater depth in subsequent interviews.

All interviews were conducted by the same investigator, an educational adviser involved in curriculum development in the faculty with little direct student interaction, but much interaction with lecturers. All interviews were conducted in a setting suggested by respondents. Interviews were conducted in either English or Afrikaans, according to respondents' preference. Care was taken to alert respondents to the fact that their personal accounts were of interest, so that they recounted their own experiences and views rather than what they may have perceived the interviewer to want to hear. Several respondents had to be encouraged to relate their personal experiences and approaches "warts and all", rather than their sanitized impressions of how they thought they should be learning or of how they perceived the nebulous "they" (i.e., other students) to approach learning and assessment. Despite being given an undertaking regarding the confidentiality of data at the start of each interview, several respondents also had to be reassured during their interview about the confidentiality of their comments, before they proceeded to share information they perceived could elicit unfavourable responses from the lecturers concerned. That said, almost all interviews "caught fire" and had to be carefully kept on track as respondents enthusiastically discussed the topic at hand.

All interviews were audio recorded and transcribed verbatim, to ultimately generate almost 1,000 pages of transcripts. Data analysis commenced even as data collection proceeded. Before progressing to more detailed analysis, field notes were reviewed and each transcript was read to obtain a global impression of how assessment impacts on student learning. Initial open coding was then undertaken by one of us (FC). As data collection and analysis progressed, codes were developed, refined and revised in an iterative process (Charmaz 2006; Dey 1993; Miles and Huberman 1994). Ongoing data collection, comparisons of codes within and between interviews and discussions between team members served to confirm and clarify codes. Clustering and partitioning of codes led to the emergence of categories as data analysis progressed, which categories were also iteratively refined, revised, discussed and ultimately related to one another.

As analysis progressed and relationships between constructs became more established, it became evident that various dimensions of motivation and emotion featured prominently when exploring the link between assessment and learning. Focussed coding of the existing dataset at that point was undertaken. However, while confirming a role for motivation and emotion, this proved to be an inadequate explanatory framework. In many instances, it was simply not possible to label a mechanism by which assessment exerted an influence on learning using this framework. Despite extensive efforts re-appraising existing data and exploring constructs in subsequent interviews, no further useful constructs could be discerned. In fact, nothing new emerged during data collection subsequent to interview fourteen, despite the individualized nature of each interview and adaptations that were made on the basis of preliminary data analysis. Analysis stalled at this point, it being apparent that a framework was needed that transcended motivation and emotion.

Recourse was had to the literature. This was informed by memo's generated during data analysis up to that point, that suggested a prominent role for two constructs i.e., imminence of assessment and the consequences of assessment. It became apparent that the variables at play were discernable as determinants of action (Bartholomew et al. 2001; Dörnyei 2000; Gebhardt and Maes 2001). At this point, focussed coding of the entire dataset was

undertaken again, this time using the various relevant constructs from this literature. This entailed some refinement of existing codes and the introduction of some new ones. However, not only were constructs relating to motivation and emotion successfully embraced by this new framework, but data that had previously proved recalcitrant to analysis also yielded to it. During this recoding process, no new constructs or relationships emerged from analysis of interviews 13–18.

Results

Where interviews were conducted in Afrikaans, quotations have been translated. Respondents will be identified as indicated in Table 1.

The impact of assessment on learning was mediated through various determinants of action. Respondents’ learning behaviour was influenced by **appraising the impact of assessment, appraising their learning response**, by their **perceptions of agency** and by **contextual factors** (Fig. 1).

Appraisal of impact

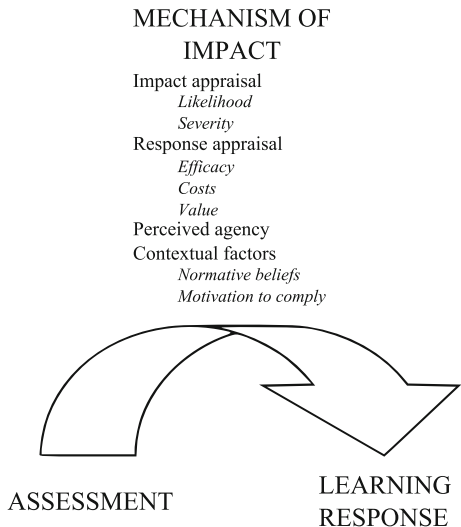
Respondents considered two factors relating to the impact of assessment: how likely consequences were to accrue and what the magnitude of consequences was likely to be.

Likelihood of impact

For these respondents, the potential for assessment to impact their lives was an unvarying certainty, inescapable, given that assessment in each module determined progression within the program.

[Quote 1] You leave things out that you think they will not ask. So it’s maybe big things or maybe important things that could save a patient’s life 1 day, but you don’t swot it because you have to pass the test now and that’s a problem for me. **(6(V)F65)**

Fig. 1 Mechanism of impact of assessment on learning



Magnitude of impact

The magnitude of any (positive or negative) impact of assessment also influenced respondents' learning. This impact could be external, on progression towards goals like passing a year (cf. Quote 8) or being certified to practice medicine (cf. Quote 3). It could be internal e.g., generating anxiety (cf. Quote 5). Interestingly, the threat of failure loomed large for many respondents, notwithstanding the fact that most of them had never failed a module (Table 1).

[Quote 2] initially, I studied more, as sick as it may sound, it was really actually nice to learn new things. Now, it's more so that I know when I go and write exams, it's just for my own peace of mind too. One learns so that you know you are not going to fail. That's a big motivation. (2(V)F77)

This cuts both ways, however. The **lack of (or limited) consequences** of other assessment e.g., in-module assignments that contributed but a small proportion of the class mark for the module, also impacted on respondents' learning. Where such assignments were used, respondents would ration their efforts and knowingly sacrifice the small number of marks on offer in the knowledge that effort spent on preparing for the end of module assessment would accrue greater rewards. This was the case even where respondents were interested in and/or enjoyed the assignments during the module. If the "reward" on offer was not great enough, they were less likely to make a concerted effort.

Appraisal of response

When contemplating assessment, respondents variously considered the **efficacy** of any given learning response in achieving a particular outcome, the **costs** of that response and the **value** of that response as measured against the respondent's personal goals and their conceptions of success and wellness. The learning response to assessment was typically not considered in isolation, but rather balanced against demands from and interests in other dimensions of respondents' lives.

Response efficacy

Respondents adapted their learning behaviour in various ways—even to the point of adopting approaches dissonant with their longer term goals (cf. Quotes 1, 8) or detrimental to the quality of their learning (cf. Quote 5)—to meet what they perceived to be the demands of assessment. In some instances, respondents matched the type of learning to the type of assessment. They reported memorizing lists for assessment purposes, even though this meant deliberately not studying for insight e.g., not learning material like pathophysiology or in ways perceived to be beneficial to patient care.

[Quote 3] I can swot lists for tests, but I forget those again. And it... it frustrates me unbelievably much if I don't have insight.

INTERVIEWER: Why do you swot lists then?

RESPONDENT: Well, I must get the marks. You know you must pass your course to become a doctor. Whether you agree with what they asked or not. (10(V)M82)

In other instances, respondents reported calibrating the magnitude and distribution of effort based on past experience or the workload relative to the time available.

[Quote 4] I think you tend to stress more earlier on the course, so you, you maybe... I think you actually relax. You know you can get away with more later on, so you maybe start studying a little later. If you have experienced before that, okay, I can actually start at this stage and still be fine, then you... then you tend do it the next time in the same way. **(4(V)F81)**

Other respondents reported matching their learning response to their perceptions of what an examiner was likely to ask. Respondents were less likely to leave work out if they perceived assessors to be less predictable e.g., if assessors did not either focus on common conditions in test papers or repeat questions from one test to another. Respondents also reported being more likely to pay attention to “spots” if they did not have enough time to review relevant work before a test.

Response costs

Costs could be incurred by responding, and not, to the demands of assessment.

Meeting the demands of assessment incurred costs internal to and external to the respondent. *Internally*, these included tension between the short term goal of success in assessment and longer term goals of delivering quality clinical care. Respondents were sceptical that success in much of the assessment they were subjected to would correlate with good quality cognitive outcomes or future success as a clinician. They reported frustration resulting from wanting to learn to be good clinicians, but having to compromise both the type of content they learned and the way they learned that content in order to meet the short term goals of passing assessment (cf. Quotes 1, 3, 8). They were, however, pragmatic about the fact that if they do not pass the assessment and thus the year, there will be little point in knowing the material anyway. Respondents thus knowingly sacrificed long term benefit (better quality knowledge and being a better doctor) for short term gain (passing the assessment).

Respondents also reported adopting strategies to reduce stress associated with assessment:

[Quote 5] ... in our third year, we could progress [to the next year], so if you got 65, then you didn't need to go and write exams. Now, I cum'd all my theory modules at the end of the day ... and so I didn't go and write the exams. And the one... there I go and get 65 on the nose. Then my parents said “are you going to do the exam”. Then I said to them “no, I've got 65. I'm going to progress”. I'm not prepared to just, so that it will stand on paper that I cum'd that thing too, now put myself through all that stress of going to learn again and write an exam again. ... to go through the work again another time is, I suppose, always advantageous, but at that time, the pros and cons were just for me ... it was just not that important to me. **(8(V)F81)**

Emotional costs accrued from **not responding to assessment**. The anxiety related to assessment was a driver of when respondents started paying increasing attention to their learning rather than other aspects of their lives. As assessment became more imminent, so respondents' anxiety levels increased, to a point where the anxiety acted as one influence on the distribution of their learning efforts (cf. Quote 4).

[Quote 6] I think if there is no assessment, you won't learn, because that is basically what it's about in medicine. You have a 4 week module, for example, and you know the first week that you must get an overview of the work and then you know when your stress mechanism starts to kick in, you must now start learning and then you start learning.

INTERVIEWER: What makes you stress?

RESPONDENT: I feel if I didn't go through all the work properly... I won't pass the test. I always feel like that. I know it's sometimes a bit in you somewhere... you only have to learn up to three quarters of the work but I feel if I didn't learn all the work and am satisfied with what I did that there is a possibility that could fail and I don't want to do that. **(9(V)F83)**

Externally, costs accrued to other assessment tasks (as described above) and to non-learning activities. Costs accrued to non-learning activities from the unrelenting nature of the workload and from the periodic peaks in workload associated with frequent end-of-module assessment. Respondents reported wanting a break (cf. Quote 5), both mentally and to allow attention to other areas of their lives, and taking this even at the cost of allowing work to accumulate beyond the point where they would be able to deal with it all prior to assessment, and consequently have to engage in compensatory tactics to ensure success.

[Quote 7] ... you get to the stage of like, when you feel you've worked hard on your last clinical block and you feel like relaxing, socializing, then you see you've got a test. The first week of the module: not a good way to start. But I mean it definitely makes you start earlier. **(4(V)F81)**

Value attached to expected outcome

Learning responses were variously geared to ensuring respondents earned the "reward" of passing assessment (cf. Quotes 1, 8) or avoided the "punishment" of failing (cf. Quotes 2, 6, 9), depending on their goal orientations. While goal orientations were not formally determined, it was clear from interview data that various respondents displayed perform approach, perform avoid and mastery approach orientations (Pintrich 2003). Some articulated more specific goals, like achieving a pass with a certain score.

Respondents reported experiencing tension between different course-related goals e.g., successfully negotiating assessment and being well prepared for patient care. As assessment loomed, the goal of passing attained sufficient value that respondents reported learning in ways that were dissonant with their goal of becoming a good clinician (cf. Quotes 1, 3).

[Quote 8] ... as far as assessment goes, I will easily go and look at an old question paper or two or so, tips that other students give and based on that, I will go... go learn, focus on certain things. And, to my own detriment for the day that clinical comes, skip some things, so then I didn't emphasize those, but when you're in a corner... when you are calmly underway, then you feel "I must just swot... I want to swot to be a good doctor", but when you are in a corner, then you swot to make one exam, because you know... Yes, the pressure is pretty high some days, because then you know that your whole year can hinge ... on this one exam. In theory, you can plug your year... so [laughs] then some days all the good intentions go out the window. **(7(IV)M72)**

However, becoming a good clinician became a more prominent goal as respondents entered their fifth year of study and the 18 month student internship, which started in semester 10 of the program, loomed.

[Quote 9] In the beginning, [my main motivation to swot] was a fear “I’m going to fail”. Crumbs, I... I firmly believed I wasn’t going to make my first year. So, it’s really just that drive to go and sit. But now it’s just... I ... especially now in the later stages, not just for myself that I ... but also for my patients that I’ll work with one day. So ... I, I ... owe it to them that ... that I must know that I must do the best that I can and I owe it to them to be a good doctor one day. **(5(V)M73)**

The value accorded assessment was thus not fixed across time and waned somewhat as students became more senior. For fifth-year students, the imminence of becoming more involved in patient care as student interns became an increasingly prominent factor in their learning. However, they still felt themselves forced to learn in ways that were at odds with achieving this goal. Thus, the looming student internship heightened the tension respondents experienced between pragmatically having to learn in ways that will help them pass exams, and idealistically learning what they believe will be useful to help them care for patients.

In addition to negotiating the tension between these course-related goals, at any given time, respondents also weighed the costs of responding, or not, to assessment against attending to the many other imperatives and interests in their lives. Respondents’ interest in, and value attached to, any one of these fluctuated constantly, generating a dynamic, ever-changing motivational mosaic. However, each time assessment loomed, the value attached to learning relative to other activities typically grew (cf. Quote 7).

[Quote 10] ... the first week, if I don’t understand something, then I go back and I’ll again... go and get the hang of it. While other swotting thingeys, I won’t in my first week. I’ll go and put in an hour or two hours each day. Then the second week, then I’ll concentrate more on my sport as a... And then in the second week, then one starts picking up a bit. There will be days, say like a Tuesday, then you’ll take off and maybe go watch a movie or so, but you pick it up a bit. In the third week, then you start about four hours or so. And then from that... you already start... the second week’s weekend, you already start putting in a bit. The first weekend, I don’t really put in. And then the third weekend, then you start, from there, you start putting in hard. **(5(V)M73)**

[Quote 11] I don’t know how else... how else one can do it, because there is just not enough time. How, even if you start swotting on the first day of the module, which never happens ... You always say to yourself “I’ve worked so hard now, so I’m going to take the first week off now”. And then... one sees one week of four has passed and then the second week starts and then you must already catch up the first week’s work. **(6(V)F65)**

Respondents’ interest in the topic of assessment also played a role in determining the amount of effort they devoted to it, particularly with small assignments during modules. The more interested they were in the topic, the more effort they were likely to devote time and effort to the task.

Perceived self-efficacy

A sense of self-efficacy has to do with the perception of being able to exert some control over a situation, even in the face of adversity. Respondents reported developing a sense over time of what they were able to achieve academically in any given time frame, and being able to calibrate the magnitude, distribution and nature of their learning efforts to achieve their predetermined goals when being assessed (cf. Quotes 4, 6).

[Quote 12] You know how long it takes you to learn something, when you must start waking up and you know when you are behind. And you know what your abilities are and how much you must do to... to be able to get to the test. **(8(V)F81)**

Various factors challenged respondents' perceived self-efficacy in relation to assessment, including unfamiliar question formats, unpredictable assessors, work overload and modules with a reputation of being difficult. This resulted in respondents adopting compensatory tactics like spotting and memorizing rather than studying for understanding.

Contextual factors

The most important contextual factors in respondents' context were various referents. Referents, people whose opinion an individual values, play an important role in influencing intent and behaviour. They provide normative beliefs against which an individual can calibrate their behaviour, if so motivated. For respondents, two key groups of referents were lecturers and other students. Lecturers served as referents both directly and indirectly. Students could include both peers of the respondent and students who previously successfully negotiated the particular year of study. Some respondents clearly fell in the category of cue seekers as described by Miller and Parlett (1974). Others were cue conscious and became more so the greater the degree of trouble they perceived themselves to be in.

Normative beliefs

Respondents actively sought out cues as to what **lecturers** believed to be important and therefore more likely to feature in assessment.

[Quote 13] one gets a reasonable idea in later years what the lecturers think is important. So, if I can, I will swot everything that there is to swot, but if the time gets a bit too little, then I take... not chances, but then I concentrate more on the things that to me are clearly more important. **(2(V)F77)**

On the one hand, respondents reported taking cues about what content to learn from the broader configuration of the curriculum. They reasoned that, given the extent of material that could potentially be covered in any given module, material that lecturers chose to focus on and include in lectures was likely to be of greater relevance to clinical practice than other material and also, therefore, more likely to feature in assessment. Within this "set" of information, they focused more on a subset of information defined by the fact that, say, three lectures were devoted to a particular topic. In this way, respondents defined for themselves what they believed to be "most assessable material".

[Quote 14] My theory is that a lecturer will not ask me something in an exam that he did not go to the trouble of mentioning in class. **(7(IV)M72)**

[Quote 15] ... if you paid attention in class, because in the 45 min, they cannot do the whole pack of notes. So, they touch on certain topics. And you know that, ten-to-one, the overwhelming majority will come out of the things they touched on in the lecture. **(8(V)F81)**

Respondents also reported attending class to glean more nuanced information about what the lecturer considered important with a view to deciding what content to focus on when learning and refining what they considered to be "most assessable material". This

went beyond direct cues like “I won’t ask this in the exam” or “Expect this in either the test or the exam”. Respondents also attended lectures to discover what material lecturers subtly emphasized while lecturing. This included taking cues from how work was presented.

[Quote 16]... obviously you have the past papers, but I mean we also have... I mean, the way that classes are presented. The professor said the whole time how important the principles are and he spent a great deal of time on these principles. You must understand, he said again and again, “listen, doses and such stuff is the type of thing you’ll be asked in your [student intern] year. Now we want you to understand the principles and understand how stuff works”. So I think it became more apparent in the way classes were presented. ... I mean, if the lecture goes on with a guy that puts up PowerPoint’s that click, click, click, click, click and here comes a bunch of information, the next slide. You know, he’s not really going to test your insights, because he didn’t try and explain the concepts to you at all. He just simply gave you facts. So you can just expect that the paper will be factual. (10(V)M82)

Respondents also sought indirect guidance from lecturers by consulting past papers (cf. Quotes 8, 16). Some respondents did this proactively i.e., at the start of their studies in a module. This was done to gather information about the type of questions asked so as to understand what type of content to focus on and not. Other respondents reported consulting past papers when they were in a pinch and they realized they could not learn all the material they had hoped or planned to. They did this so as to adapt their learning plans.

In some cases, this strategy allowed respondents to make definite choices about including or excluding content from their learning. In some cases, no guidance could be obtained from lecturers or past papers to focus studies. No “spots” were evident. For cue-seeking students, this resulted in high (though typically not debilitating) levels of anxiety.

Many respondents also sought or took cues and guidance from fellow **students** (cf. Quote 8) both in their class and from more senior classes, even acknowledging that the reliability of information from this source varied. In some instances, this was proactive and related to the workload or degree of difficulty of the work and therefore the amount and nature of effort required in and out of class to succeed in assessment. In other instances, it related to identifying work to focus on and leave out, both ahead of time and in the final hours before an assessment.

Motivation to comply with normative beliefs

Many respondents’ motivation to comply with normative beliefs increased proportionally to their perception of the decreasing likelihood of success in assessment (cf. Quote 8). As mentioned above, it was also evident that some respondents’ motivation to comply with the perceived beliefs of lecturers or fellow students lessened somewhat as the student internship became more imminent.

Negative cases

As part of the process of analysis, data were scrutinised for evidence of cases which gave evidence that this mechanism was not a valid explanation of the link between assessment and learning. There were cases where students clearly were strongly focussed on their upcoming professions. The fifth-year respondents clearly were far more aware of learning for the purposes of caring for patients than their fourth year counterparts. The imminence

of their student internship played a great role in this regard. Some respondents also seemed to have a more professional orientation in general or to be learning for the love of learning. However, it was striking how modulating factors like the imminence of assessment and the prevailing workload rode roughshod over such orientations. Even in the respondents with the most strongly established learning orientations of this nature, there was clear evidence that imminence, as a modulating factor, overrode their intrinsic orientation and that elements of the proposed mechanism came into play. Impact appraisal resulted in their abandoning their preferred learning activities in favour of strategies like, for example, memorising lists that would bring success in assessment, which success held the key to their progressing to the next stage of the course and being able to provide care for patients, their actual goal (as opposed to merely passing the exam). They also abandoned learning material they believed was relevant to patient care in favour of learning material they believed was irrelevant to patient care but relevant to achieving success in the upcoming assessment. Ultimately, then, there were no cases where some dimensions of this mechanism were not at play or where an alternate mechanism could be identified as the dominant factor at play.

Discussion

The purpose of this exploratory study was to probe the mechanisms by which assessment impacts on learning, focusing on how various dimensions of summative assessment of theory bring about such influence as they exert on various dimensions of learning. What is proposed here is a theoretical framework to explain why students respond to assessment in the ways that they do and not merely what their response is.

The factors described in this study do not form a simple target for intervention. Not all of these factors are in play for any given student at any given time, nor are the factors in play for any given student constant across time and context. Even if the same factors are in play for two students, the intensity of that impact may vary based on personal or other contextual influences individual to each. This serves to underline that “the social and cultural contexts of the phenomenon studied are crucial for understanding the operation of causal mechanisms” (Maxwell 2004a: 6). To complicate matters even further, as Gebhardt and Maes (2001) caution in the context of health behaviour, not all behaviour is the result of a considered response to the factors inducing the behaviour. Hence, when assessment is manipulated to influence learning, students may initially act out established patterns of behaviour rather than making the effort to make considered, deliberate changes. This might go part way to explaining the lack of desirable impact of thoughtfully designed assessment interventions.

It also bears emphasising that assessment is typically not a single, homogeneous entity to which students respond. Each module or course makes its own demands of students, often independently of rather than in synchronised fashion with others. Ultimately, to have an impact on as many dimensions of learning of as many students as possible, it will be necessary to manipulate multiple dimensions of assessment i.e., assessment systems, rather than tweak individual assessment events. Yet not taking this complexity into account will result in failed efforts to positively influence learning using assessment. That said, what do these results suggest about designing assessment to enhance quality learning?

Assessment that is intended to impact learning should have consequences, be those consequences on students’ marks and progression or on other factors e.g., their esteem in the eyes of fellow students as may be the case with project presentations or peer assessment

or their sense of agency as may be the case with feedback. Furthermore, any attempt to manipulate assessment to influence learning has to be considered against the backdrop of all assessment in the system—particularly high stakes, summative assessment—that will also be impacting on students' learning.

The degree of impact of any assessment activity will probably be strongly correlated with the magnitude of the severity of consequences associated with it. Introducing feedback on a one-of-a-kind assignment contributing 10% of a student's grade is likely to be a less successful intervention than changing the level of cognitive challenge from recall to problem-solving in a multiple choice assessment contributing 50% of a student's grade.

Assessment should be designed in such a way that when students make their appraisal of the efficacy of their learning response, that appraisal leads them to learn in ways we as academics believe they should be learning. Evidently, Newble and Jaeger (1983) and Frederiksen (1984) were able to bring about just such changes. In theory modules, using a mixed bag of longer and shorter questions types would force students to engage differently with learning material, both in terms of what they learn and in terms of how they learn. They would be unable to simply leave out work based on the length or type of questions to be asked. Having one longer (e.g., 20 mark) question per assessment mixed with shorter questions should lead to qualitative differences in students learning. Clearly, such a design would also have to be weighed against various other, including psychometric and pragmatic, considerations (van der Vleuten 1996).

The costs to students of any given learning response should not be too high. Where the demands of an assessment system become too onerous, the cost-benefit analysis will lead students to find short cuts. One thinks here of the introduction of portfolios for assessment. This is based on sound educational grounds, but does not always lead to a salubrious impact on learning (Driessen et al. 2007).

Equally, assessment tasks that challenge students' sense of agency by virtue of being unknown or complex, or based on material too complex (van Etten et al. 1997) or too voluminous to engage with meaningfully in the limited time available within an academic module, will be unlikely to have a positive impact on student learning.

Finally, the impact of the "myths and legends" about assessment that swirl around a module and/or a lecturer should not be lost from sight. The university where no volumes of past examination papers are passed from one generation of students to the next is likely a rare place indeed. These form the basis of analyses of content and style that inform the learning of many a student. As such, they should be brought out of the shadows and incorporated openly into the assessment system.

There are some limitations inherent in this work, largely related to the methodology employed. One of the innate dangers when utilizing in-depth interviews with individual respondents is that respondents cannot be assumed to have complete knowledge about the impact of assessment on learning, or that their accounts was unbiased by their interpretation of their experiences or the situations to which they refer, or the situation of the interview (Cohen et al. 2000). However, care was taken to get respondents to describe and explore their actual responses in various situations they had experienced over time and to minimise the impact of the situation of the interview. Furthermore, data saturation was reached both during interviews and during data analysis and we are confident that the constructs identified are exhaustive for this group of respondents.

A second limitation is that all respondents volunteered to participate in the study. In keeping with other studies, there was a greater proportion of respondents with higher average scores than the classes from which they were drawn (Callahan et al. 2007; Entwistle and Entwistle 1991). This could have resulted in a greater proportion of

insightful comments, but possibly over-represented well organised study strategies. In contrast to Callahan et al.'s findings, our respondents included a greater proportion of women. Given that gender may correlate with approaches to study (Duff 2002), further study of the role of gender in this mechanism is warranted. Overall, however, as the purpose of this study is the elucidation of a process rather than generalizability, neither of these factors is considered a drawback.

How credible are our findings as a causal explanation? Our study has certainly addressed many of the threats to causal inference identified by Maxwell (2004b). This framework is derived from the lived experiences of students in a complex, authentic educational system over time. Our observation of causal processes is admittedly indirect by way of interviews. Nonetheless, even though interviews were conducted at one point in time, we explored the responses of respondents to multiple different assessment situations across their years of study. Similarities and differences between constructs and relationships between constructs were recognisable across contexts for given respondents and across respondents. Echoes of these constructs and relationships are also discernable in published literature (see below). As noted earlier, we were unable to find discrepant data or negative cases in our data that would challenge the proposed framework. Finally, we have informally solicited responses to this data from audiences in the health sciences and other disciplines in South African and international settings (e.g., Cilliers et al. 2008, 2009). Participants have invariably acknowledged the findings to represent a “recognisable reality”, both from the perspective of their own experience as learners and that as lecturers.

Does this mechanism have currency beyond the setting in which it has been elucidated i.e., an environment that uses high stakes assessment, repeatedly, as a means of advancing medical students at one South African university through the curriculum? Clearly the research design does not hold generalisability as it's intent. Nonetheless, there are some tempting clues that we believe indicate that exploring this mechanism in other settings is warranted. When interpreted in light of this mechanism, findings reported in other literature generated from work done in various non-medical contexts in Scotland and the United States (Becker et al. 1968; Frederiksen 1984; Miller and Parlett 1974; Snyder 1971) give evidence of the mechanism being in play in other contexts. In fact, in contrast to the setting of this present work where the curriculum had been designed specifically so that students only studied one module at any given time, these other reports evidence how the mechanism plays out in settings where students have to juggle demands from multiple courses concurrently.

It is also anticipated that this model might be useful beyond the context of summative assessment. One of the conundrums with using feedback is that whilst it has been found to potentially have a powerful impact on student learning (Black and Wiliam 1998), it is often missed or misunderstood by students (see Gibbs and Simpson 2004 for examples). It is tempting to speculate that applying the lens of impact appraisal and response appraisal to some of the findings of research on feedback could shed some light on these phenomena. It is interesting to note that feedback from lecturers did not feature as a factor influencing learning in this study.

As Ramsden (1992) pointed out: “Unsuitable assessment methods impose irresistible pressures on a student to take the wrong approaches to learning tasks”. With a better understanding of how assessment impacts on student learning, it will hopefully be possible to start exploring how assessment can be better utilized to bring about meaningful student learning and remedy this situation. Crucial links in ensuring that assessment is utilized more effectively will include the academics who assess students, and administrators who increasingly decide on the mix of demands—assessment-related and otherwise—to which

academics should be answerable. Understanding these cogs in the greater academic machine will hopefully ensure that we are not still lamenting the deplorable impact of assessment on learning some decades hence.

Acknowledgments We would like to thank the students who participated in this study. This material is partly based upon work financially supported by the National Research Foundation in South Africa. Financial support from the Foundation for the Advancement of International Medical Education and Research is also gratefully acknowledged.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Airasian, P. W. (1988). Measurement driven instruction: A closer look. *Educational Measurement: Issues and Practice*, 7, 6–11.
- Alderson, J. C., & Banerjee, J. (2001). Language testing and assessment (Part 1). *Language Teaching*, 34, 213–236.
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115–129.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13, 257–279.
- Bartholomew, L. K., Parcel, G. S., Kok, G. & Gottlieb, N. H. (2001). Chapter 6. Selecting theory based intervention methods and practical strategies. In: *Intervention mapping: Designing theory-based health promotion programmes*. Mountain View, California: Mayfield Publishing Company
- Becker, H. S., Geer, B., & Hughes, E. C. (1968). *Making the grade: The academic side of college life*. New York: Wiley.
- Becker, H. S., Geer, B., Hughes, E. C., & Strauss, A. (1961). *Boys in white. Student culture in medical school*. Chicago: University of Chicago Press.
- Biggs, J. B. (1987). *Student approaches to learning and studying*. (Hawthorn, Australia: Australian Council for Educational Research Ltd).
- Biggs, J. B. (1996). Assessing learning quality: Reconciling institutional, staff and educational demands. *Assessment & Evaluation in Higher Education*, 21, 5–16.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5, 7–74.
- Boud, D. (1995). Assessment and learning: Contradictory or complementary? In P. Knight (Ed.), *Assessment for learning in higher education* (pp. 35–48). London: Kogan Page.
- Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment & Evaluation in Higher Education*, 24, 413–426.
- Broekkamp, H., & van Hout-Wolters, B. H. A. M. (2007). Students' adaptation of study strategies when preparing for classroom tests. *Educational Psychology Review*, 19, 401–428.
- Bunge, M. (2004). How does it work? The search for explanatory mechanisms. *Philosophy of the Social Sciences*, 34, 182–210.
- Callahan, C. A., Hojat, M., & Gonnella, J. S. (2007). Volunteer bias in medical education research: An empirical study of over three decades of longitudinal data. *Medical Education*, 41, 746–753.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. London: Sage Publications Ltd.
- Cilliers, F. J., Schuwirth, L. W. T., Adendorff, H. J., Bitzer, E. M., Herman, N. & van der Vleuten, C. P. M. (2008). *The impact of summative assessment of theory on how medical students learn*. Paper presented at AMEE Conference, Prague, Czech Republic
- Cilliers, F. J., Schuwirth, L. W. T., Adendorff, H. J., Herman, N., & van der Vleuten, C. P. M. (2009). *Assessment impacts on learning, you say? Please explain how. The impact of summative assessment of theory on how medical students learn*. Paper presented at Annual Conference of the Society for Research into Higher Education, Celtic Manor, Wales
- Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education*. London: RoutledgeFalmer.
- de Corte, E. (2007). Learning from instruction: the case of mathematics. *Learning Inquiry*, 1, 19–30.

- de Corte, E., Verschaffel, L., Entwistle, N. J., & van Merriënboer, J. J. G. (2003). *Powerful learning environments: unravelling basic components and dimensions*. Oxford: Pergamon.
- Dey, I. (1993). *Qualitative data analysis*. Routledge, USA and Canada
- DiCicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interview. *Medical Education*, *40*, 314–321.
- Dörnyei, Z. (2000). Motivation in action: Towards a process-oriented conceptualisation of student motivation. *British Journal of Educational Psychology*, *70*, 519–538.
- Driessen, E., van Tartwijk, J., van der Vleuten, C. P. M., & Wass, V. (2007). Portfolios in medical education: Why do they meet with mixed success? A systematic review. *Medical Education*, *41*, 1224–1233.
- Duff, A. (2002). Approaches to learning: Factor invariance across gender. *Personality and Individual Differences*, *33*, 997–1010.
- Elton, L. (1987). *Teaching in higher education: Appraisal and training*. London: Kogan Page.
- Elton, L. R. B., & Laurillard, D. M. (1979). Trends in research on student learning. *Studies in Higher Education*, *4*, 87–102.
- Entwistle, N. J., & Entwistle, A. C. (1991). Contrasting forms of understanding for degree examinations: The student experience and its implications. *Higher Education*, *22*, 205–227.
- Entwistle, A. C., & Entwistle, N. J. (1992). Experiences of understanding in revising for degree examinations. *Learning and Instruction*, *2*, 1–22.
- Entwistle, N. J., & Entwistle, A. C. (2005). Revision and the experience of understanding. In F. Marton, et al. (Eds.), *The experience of learning: Implications for teaching and studying in higher education* (pp. 145–155). Edinburgh: University of Edinburgh, Centre for Teaching, Learning and Assessment.
- Fransson, A. (1977). Qualitative differences in learning. 4. Effects of intrinsic motivation and extrinsic test anxiety on process and outcome. *British Journal of Educational Psychology*, *47*, 244–257.
- Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist*, *39*, 193–202.
- Gebhardt, W. A., & Maes, S. (2001). Integrating social-psychological frameworks for health behavior research. *American Journal of Health Behavior*, *25*, 528–536.
- Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education, 2004–05*, 3–31.
- Gijbels, D., Coertjens, L., Vanthournout, G., Struyf, E., & Van Petegem, P. (2009). Changing students' approaches to learning: A two-year study within a university teacher training course. *Educational Studies*, *35*, 503–513.
- Glover, J. A. (1989). The “testing” phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, *81*, 392–399.
- Janssens, S., Boes, W., & Wante, D. (2002). Portfolio: Een instrument voor toetsing en begeleiding. In F. J. R. C. Dochy, et al. (Eds.), *Assessment in onderwijs* (pp. 203–224). Utrecht: LEMMA.
- Jones, H. E. (1923). Experimental studies of college teaching: The effect of examination on the permanence of learning. *Archives of Psychology*, *10*, 1–70.
- Joughin, G. (2007). Student conceptions of oral presentations. *Studies in Higher Education*, *32*, 323–336.
- Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. Thousand Oaks, California: Sage Publications, Inc).
- Laurillard, D. (1979). The processes of student learning. *Higher Education*, *8*, 395–409.
- Lundeberg, M. A., & Fox, P. W. (1991). Do laboratory findings on test expectancy generalize to classroom outcomes? *Review of Educational Research*, *61*, 94–106.
- Madaus, G. F. (1988). The distortion of teaching and testing: High-stakes testing and instruction. *Peabody Journal of Education*, *65*, 29–46.
- Maxwell, J. A. (2004a). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher*, *33*, 3–11.
- Maxwell, J. A. (2004b). Using qualitative methods for causal explanation. *Field Methods*, *16*, 243–264.
- Meyer, G. (1936). The effect of recall and recognition on the examination set in classroom situations. *Journal of Educational Psychology*, *27*, 81–99.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks: Sage Publications, Inc.
- Miller, C. M. L., & Parlett, M. (1974). *Up to the mark: A study of the examination game*. London: Society for Research into Higher Education.
- Newble, D. I., & Jaeger, K. (1983). The effect of assessment and examinations on the learning of medical students. *Medical Education*, *17*, 165–171.
- Parlett, M. (1969). Undergraduate teaching observed. *Nature*, *223*, 1102–1104.

- Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*, 667–696.
- Ramsden, P. (1984). The context of learning. In F. Marton, et al. (Eds.), *The experience of learning* (pp. 144–164). Edinburgh: Scottish Academic Press.
- Ramsden, P. (1992). *Learning to teach in higher education*. London: Routledge.
- Ramsden, P. (2005). The context of learning in academic departments. In F. Marton, et al. (Eds.), *The experience of learning: Implications for teaching and studying in higher education* (pp. 198–216). Edinburgh: University of Edinburgh, Centre for Teaching, Learning and Assessment.
- Roediger, H. L. III, & Karpicke, J. D. (2006). Test-enhanced learning. *Psychological Science, 17*, 249–255.
- Ross, M. E., Green, S., Salisbury-Glennon, J. D., & Tollefson, N. (2006). College students' study strategies as a function of testing: An investigation into metacognitive self-regulation. *Innovative Higher Education, 30*, 361–375.
- Ross, M. E., Salisbury-Glennon, J. D., Guarino, A., Reed, C. J., & Marshall, M. (2003). Situated self-regulation: Modeling the interrelationships among instruction, assessment, learning strategies and academic performance. *Educational Research and Evaluation, 9*, 189–209.
- Sambell, K., & McDowell, L. (1998). The construction of the hidden curriculum: Messages and meanings in the assessment of student learning. *Assessment & Evaluation in Higher Education, 23*, 391–402.
- Scouller, K. M. (1998). The influence of assessment method on students' learning approaches: Multiple choice examinations versus assignment essay. *Higher Education, 35*, 453–472.
- Segers, M., & Dochy, F. J. R. C. (2006). Introduction Enhancing student learning through assessment: Alignment between levels of assessment and different effects on learning. *Studies in Educational Evaluation, 32*, 171–179.
- Snyder, B. R. (1971). *The hidden curriculum*. New York: Alfred A. Knopf.
- Struyven, K., Dochy, F. J. R. C., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education, 30*(4), 325–341.
- Tang, K. C. C. (1994). Effects of modes of assessment on students' preparation strategies. In G. Gibbs (Ed.), *Improving student learning—theory and practice* (pp. 151–170). Oxford: Oxford Centre for Staff Development, Oxford Brookes University.
- Thomson, K., & Falchikov, N. (1998). 'Full on until the sun comes out': The effects of assessment on student approaches to studying. *Assessment & Evaluation in Higher Education, 23*, 379–390.
- van der Vleuten, C. P. M. (1996). The assessment of professional competence: Developments, research and practical implications. *Advances in Health Sciences Education, 1*, 41–67.
- van Etten, S., Freebern, G., & Pressley, M. (1997). College students' beliefs about exam preparation. *Contemporary Educational Psychology, 22*, 192–212.
- Vermunt, J. D. (1996). Metacognitive, cognitive and affective aspects of learning styles and strategies: A phenomenographic analysis. *Higher Education, 31*, 25–50.