Variability of Students' Responses to Assessment Activities: The Influence of Achievement Levels

Dennis Alonzo
Lecturer, School of Education, University of New South Wales Sydney, Australia, *d.alonzo@unsw.edu.au*

Tony Loughland
Assoc. Prof., School of Education, University of New South Wales Sydney, Australia, *tony.loughland@unsw.edu.au*

The effect of assessment on student learning is often reported using performance data from the entire cohort rather than the growth of individual students. This practice is inconsistent with the theoretical and empirical evidence that individual students respond differently to assessment strategies. The variance observed in students' responses to assessment tasks is commonly attributed to their learning needs and characteristics, but little is understood about the influence of their achievement levels. This study explores how secondary English students from different achievement levels respond to different assessment activities aimed at developing their skills in writing a persuasive essay and how these responses influence their further engagement in learning. An interpretivist approach was used to analyse the transcripts of semi-structured interviews conducted with five high performing, three average and three underperforming students. The findings revealed that students' responses are influenced by their learning goals and their perceived benefits of assessment strategies. High performing students selectively engage in assessment activities that best improve their learning, whilst underachieving students disengage when the activity overwhelms them. The average performing students engage only to a certain extent to meet only the average expectation or performance required. Interestingly, some of these responses did not reflect the teacher's intent in using the assessment strategy. The implications of this study suggest that effective teacher assessment practices would benefit from a recognition of the concept of stimulus-response compatibility. An adaptive teacher disposition is critical in the provision of appropriate stimuli as well as a constructive response to students to ensure their ongoing learning engagement.

Keywords: teacher adaptive practice, stimulus-response compatibility, student achievement level, assessment for learning

INTRODUCTION

Assessment practices that are grouped under the conceptual framework of Assessment *for* Learning (A*f*L) have been widely claimed to improve student outcomes (Wiliam, 2017; Alonzo, 2016). It is extensively reported that A*f*L activities that help students understand learning outcomes (Clark, 2014), clarify outcomes by using exemplars (Handley et al., 2013), provide timely and effective feedback (Hattie & Timperley, 2007) and monitor their learning progress (Rust et al., 2003) significantly improve student outcomes. However, there is also evidence that the implementation of A*f*L activities does not always influence student learning (Greve et al., 2018). Such conflicting evidence may be because most of these studies report the overall improvement of the cohort without explicitly considering the effects of discrete assessment practices on individual students. Studying the effects of assessment on individual students is more aligned with the learner-centred approach of A*f*L (Baird et al., 2017), where the focus is on helping individual students achieve better.

This study explored how students from different achievement levels engage in various assessment tasks and how their responses influence their subsequent engagement in learning in a senior secondary English classroom. Recent research articles imply the association between student achievement levels and the effectiveness of assessment activities. Most students selectively engage with assessment activities that suit their personal goals only, whilst other students do not engage at all (Colvin et al., 2016). There is also evidence that underachieving students tend to overestimate their performance when they engage in self-assessment of various assessment tasks, including written, oral, analysis and projects (Boud et al., 2015). Furthermore, in the higher education context, underachieving students want more explanations of the learning outcomes, an exercise that high achieving students find unnecessary (Gynnild, 2003). However, this implied evidence is yet to be supported and established in the case of English classrooms. This unexplored area on the association between student achievement levels and their engagement in assessment presents an opportunity for further research. Hence, our study will determine if students' achievement levels influence their engagement in assessment. We will provide empirical evidence to what has been implied in a few reported studies that student engagement in assessment is partly dependent on their achievement levels. Similarly, our study will contribute to understanding why teachers' intent in using assessment is not always translated into actual students' learning improvement.

The findings of our study will contribute to the theorisation of student-centred A*f*L practices by understanding the influence of students' achievement levels on their responses to assessment. This understanding could provide input for teachers' decision-making relative to ensuring effective implementation of assessment activities, as teachers' selection of appropriate assessment activities requires a complex cognitive process (Fives & Barnes, 2020).

Literature Review

The relationship of a student-centred approach in assessment to teacher practices

Students' experience of assessment is often reported as a generic category applicable to all students. For example, research on the benefits of engaging students in peer assessment has claimed an increase in overall outcomes, but variance in individual students' learning gains was not addressed (Billany & Billany, 2009). A fine-grained examination at the individual student level is warranted as the A*f*L framework is informed by a student-centred approach to assessment (Baird et al., 2017). The practice of A*f*L requires teachers to use their adaptive expertise, requiring teachers to strategically adjust their assessment practices depending on students' learning needs, characteristics, support needed and sociocultural background (Allen et al., 2013; Loughland & Alonzo, 2019) to develop and implement differentiated assessment activities. Differentiation in assessment, 'an educational structure that seeks to address differences among students by providing flexibility in the levels of knowledge acquisition, skills development and types of assessment items undertaken by students' (Varsavsky & Rayner, 2013, p. 790), is effective in raising student outcomes (Ginja & Chen, 2020; Magableh & Abdullah, 2020). This adaptive expertise in assessment was implied in Hattie's categorisation of expert teachers as those who are "more adept at monitoring student problems and assessing their level of understanding and progress, and they provide much [more] relevant, useful feedback" (2003, p.7). This characteristic of teachers is critical for developing an adaptive disposition associated with effective teaching (Wetzel et al., 2015).

Student perspectives on assessment

There is growing evidence that students are receptive to how teachers use assessment in the classroom. For example, Cowie (2009) found that teacher assessment practices that support learning and create a better classroom environment are valued by primary students in various key learning areas. They react to it positively by engaging more in their learning. Similarly, students who perceive assessment as a process that makes them accountable for their learning have higher achievement, whilst lower achievement is observed in students who perceive that assessment interferes with their learning (Brown & Hirschfeld, 2007). This variance in perception is reflected in students' participation in assessment activities.

In a study of peer assessment using clickers, there were undergraduate students who conscientiously completed all peer assessment activities in arts courses whilst others just randomly clicked the numbers (Barwell & Walker, 2009). Those students who randomly assigned marks viewed engagement in peer assessment as an extra work. Another study shows that much of the peer feedback on writing tasks was ignored by undergraduate students (Walker, 2015). Little of this evidence of the variance in student responses on assessment has been used by teachers to modify their assessment practices (Brown et al., 2009). This oversight has the potential to create a mismatch between teachers' intention in using an assessment strategy and their students' response to it. The variance of

student responses reported in the literature suggests that it is unlikely that one assessment strategy will work for all students.

Alignment and misalignment between teacher's intention and student response

A student's response to any assessment activity depends on how they view it (Cowie, 2009), and this view may not align with the teacher's intention (Brown et al., 2009; Orsmond & Merry, 2011; Remesal, 2009). As a result, distinct groups of students in one class will respond differently to the same assessment activity. Hence, it is not logical to expect that an assessment strategy will have the same effect on increasing outcomes for all students. This notion is supported by research that found that undergraduate students have individual preferences in terms of assessment activities, and they react differently in following their preference (Healy et al. (2014). There is no homogenous preference and reaction across all students, so even the most credible assessment activity will result in the same outcomes for all students.

There is a growing evidence base on the variability of student beliefs, views, and perspectives in assessment (Lew et al., 2010; Remesal, 2009), which is the same as teachers' assessment beliefs that are varied due to their prior knowledge, experience, and dispositions (Alonzo et al., 2020). However, little is known about the source of these variations. Student learning needs and backgrounds are often theorised to be the factors, but only gender has been found to influence student engagement in assessment. It has been found with undergraduate students that in self-assessment, females rate themselves lower than their actual performance (Torres-Guijarro & Bengoechea, 2017).

Another example of the disjunction between teacher intention and student response can be found in the work of Gao (2009). He found that formative assessment was not valued by Hong Kong primary and high school students whose perception of assessment was more akin to accountability than improvement. These students responded less to A*f*L activities due to their perception that these would not affect their final mark. This is consistent with the study of Healy et al. (2014) who found that students who viewed assessment as a reward for their effort prefer summative assessment whilst those who view assessment as valuable to their learning prefer on-going assessment.

A disjunction between teacher intention and student response can be found also in teachers' use of feedback. The nature of feedback given by teachers impacts upon student learning differently. The feedback that provides scaffolding and praise could develop students' self-regulation whereas verification feedback and directive feedback can reduce students' use of organisational strategies (Guo & Wei, 2019). The way students respond to feedback also influences the overall impact of feedback on their learning (Skovholt, 2018).

Student achievement levels as a source of variance

Student achievement levels are one possible source of the variance, but this has not been explicitly addressed in the literature. Implicit evidence of the effect of achievement levels on students' perception of assessment can be gathered from other studies. Boud et al. (2015) found that undergraduate students who are underachieving to overestimate

their performance, and they show no significant improvement in calibrating their judgment even with repeated self-assessment and feedback. In addition, Gynnild (2003) found that underachieving students expressed the need for more explicit learning outcomes to understand what is required. More evidence is found in the paper of Colvin et al. (2016) where high achieving students engaged only in assessment materials that aligned with their personal goals. There are also cases of undergraduate students where rubrics were introduced to guide their work, and only the high achieving students showed improvement in their learning (Kite & Phongsavan, 2017). Another example of an association between students' achievement level and response to assessment is how students receive and act on feedback. There is a big difference between how high and underachieving students self-assess and self-regulate when receiving feedback, which is essential for the effectiveness of how they use feedback to improve their work (Orsmond & Merry, 2009). In fact, a differentiated approach to feedback based on student achievement level has been proposed by Shute (2008), who argues that teachers need to tailor feedback content, timing, and amount to students' level of performance.

Theoretical Framework

The variability of student responses to teacher's intentions in using an assessment strategy can be theorised from multiple perspectives, from the psychological through the sociocultural to the critical, but in this study, we drew on a behaviourist approach to explain student responses using the concept of stimulus-response (SR) compatibility (Haug & Ødegaard, 2015). This theory proposes that in a biological system, every stimulus perceived by an organism draws a response due to its inherent characteristic of irritability. If the stimulus is within the threshold of what an organism can respond to, it will produce a positive response to adjust its internal system to thrive in that environment. However, if the stimulus is too much to what the organism can tolerate, it will disrupt its biological system, causing some abnormalities or even death (Kilgour, 1987). The SR compatibility is also applied to human behaviour, as Kornblum et al. (1990) postulate that people engage in independent response selection depending on the attributes of the stimulus. The response is positive and automatically processed if a person's perception of the stimulus linked to the task requirements is compatible with the action required. The stimulus-response translation and execution are automatic and interwoven because they share common level processing (Hommel, 2009; Richez et al., 2016). This means that higher compatibility between a person's perception of the task and the stimulus would bring a positive response. If the person perceives the task given to be beyond their capacity to respond to it, they are more likely to disengage from performing the task.

Applying the SR compatibility theory in the classroom, teachers need to selectively use assessment activities that will trigger positive responses from students to improve their learning. This process ensures that student perception of an assessment activity (stimulus) is compatible with the intended action (response). The basic tenet of the stimulus-response theory is that the stimulus should be related to the person's perception of the phenomenon for the required action to be activated and performed (Miles & Proctor, 2009). Otherwise, the response will be slow or different from expected

(Kornblum et al., 1990). In the case of student learning, teachers' approaches to helping students directly impact their achievement (Lau & Ho, 2016). This theoretical lens may be a generative one to understand the efficacy of teacher assessment practices, which could potentially inform teacher professional development in assessment (Alonzo, 2020).

METHOD

To achieve the aims of this study, an interpretivist case study research design was used (Neuman, 2013). To fully understand the experiences and perceptions of research participants, multiple data sources were used. It is a common characteristic of case study research to use multiple data sources for triangulation and to address issues related to validity and credibility (Eisenhardt, 1989; Yin, 2009). A combination of document analysis, classroom observations, and semi-structured interviews (Bryman, 2016) were used (See Table 1) to gather significant information on how students from different achievement levels respond to various assessment activities used in in a senior English classroom in the Philippines, where all students are studying across the curriculum through the medium of English. The assessment practices of an English teacher teaching a Year 11 cohort of 25 students (age range 16-18; males = 9 and females = 16) were observed for five one-hour sessions to understand how the teacher implemented various assessment activities in the classroom.

Table 1
Case study data generation and analysis phases

Phase	Data Source	Analytical Approach
Document analysis	Research articles; Reports	Qualitative content analysis utilising NVivo to establish relevant assessment practices of teachers in developing assessment activities, practices, processes, and other relevant dimensions. Inductive qualitative analysis was employed in this phase.
Observation	Class observation	Thematic analysis from observation notes and memos.
Interview	Semi-structured interview	Readings of raw data to derive from pre-determined concepts and open coding. A combination of deductive and inductive thematic analysis using four themes identified in phase 1. Data were coded using NVivo that revealed similar patterns consistent with the pre-determined themes.

The unit of work implemented for six weeks aimed to develop students' ability in persuasive writing. All learning, teaching, and assessment activities were designed, leading students to write a persuasive essay on a particular issue they chose. After submitting their outputs, all students were invited for semi-structured interviews. Eleven students (males = 5; and females = 6) responded to the invitation with parental consent: five belonging to the top rank of the class, three from the average rank, and three from the bottom rank. The ranking of the students was determined using the cumulative achievement tests of students in Terms 1 and 2. The classification of students based on their achievement level was not communicated to students but only used for research purposes.

The individual interview was conducted in the students' mother tongue using guiding questions, which asked about their responses to the assessment activities and their suggestions for further improvement of their teacher's assessment practices. Also, the teacher was interviewed to determine her intent in using the assessment strategies to support students. The interview guides were checked by two assessment experts who have track records in research, to establish the alignment of questions to the aims of the study. Their feedback, including explicit wordings and exclusion of irrelevant questions, were used to revise the guide questions. Pilot interviews were conducted with two students and one teacher to test the questions' appropriateness, language's suitability, and viability of research. Based on the first author's experience and the results of the pilot interview, the guide questions could provide the data needed to address the aims of this paper.

Copies of the original transcript were given back to students for member checking, then translated into English, and two research assistants verified the translation. All collected data were digitally recorded, transcribed, and stored in NVivo for easy retrieval and processing. Thematic analysis was conducted following principles of interpretivist analysis involving an inductive and iterative process. The data was broken down into recurring themes and then regrouped into coherent picture to achieve the aim of this paper. Students' responses were coded based on their alignment to teachers' intent of each assessment strategy. Students' responses were coded based on their alignment to teachers' intent of each assessment strategy. The thematic analysis involved three steps. In the first step, the lead author read the transcript and assigned codes to the data and highlighted quotes relevant to the paper's aim. In the second step, analysis was carried out in Nvivo for easy coding, decoding and recoding of data. In this step, the codes generated were aligned to the four most common assessment activities used, including sharing learning outcomes, using exemplars, eliciting and giving feedback, and engaging in self and peer assessment (Black et al., 2006). In the final step, each theme was further broken down into more specific categories highlighting the responses of underachieving, average and high performing students.

FINDINGS

The study compares students' responses from the three different performance levels (High Achieving = HA, Average Achieving = AA; and Under Achieving = UA) with that of teacher intent in using the four most common assessment activities. The findings will now be reported. The summary of the results is shown in Table 2.

Based on Table 1, the teacher's intent for using any of the four assessment strategies is not often translated into actual outcomes because students respond to them differently. For example, the teacher shared learning outcomes to remind students of the targets and use them to guide their learning, but other students find this process boring and a waste of time. Similarly, the teacher required all students to analyse a range of exemplars to identify insights they could use to improve their work. However, some students were overwhelmed with the best exemplars, while others found the "bad" exemplars useless. The subsequent subsections explicitly present the association between student ability levels and their responses to assessment activities.

Table 2

Summary of teacher intent and student responses

Teacher Intent	Underperforming Students	Average Performing	High-Performing
Sharing Learning Outcomes • Remind students everyday about the learning outcomes • Use LOs as guides • Use as checklists	• Help focus learning goals • Use as checklists • Guide for determining which one need to be learned more	• Easy learning outcomes need not be repeated • Find the learning outcomes helpful to guide in studying • Need help to the more difficult ones	• Simple explanation is enough • Find it boring if teacher keeps repeating it
Using Exemplars • Require all students to analyse a range of exemplars • Students will learn from different exemplars • They can align their targets to these exemplars	• Overwhelmed to see the "best" exemplars • Developed self-doubt upon seeing the highest exemplar • Want to see annotated exemplars within their ability level	• Sometimes look at the "best exemplars" but mostly analysed the mid-range • Ignored the low-level exemplars	• Analysed only the "best exemplars" • Found the low-level exemplars useless/waste of time • Want to see best exemplars with key points on how to exceed these exemplars
Eliciting and Giving Feedback • Provide feedback for individual students • Guide students to act on feedback • Identify as many as possible key areas for improvement	• Appreciate the feedback given but selectively act on them • Overwhelmed if there are so many feedback; make them self-conscious • Prefer to receive explicit suggestions on how to improve their work • Want to see more of feedback related to their strengths to motivate them	• Act on some feedback depending on their capability to address each • Want to receive a reasonable number of feedback • Ignore the written feedback if happy with the mark • Feedback should be given prior to submitting the final draft	• Ignore the written feedback if the mark is high • Want to receive as many feedback as possible • Act on every feedback given when asked to revise the work • Find every feedback to be useful
Engaging in Self and Peer Assessment • Force students to engage in reflective practice • They can learn from their peers	• Find self-assessment to be a daunting process • Learned from the feedback of "brighter" students • Uneasy to give feedback to high-performing peers	• Prefer to have a "brighter" student • Doubt the feedback given by a low-performing student • Confident to give feedback to a underperforming student but uncomfortable with "brighter" students	• Highly engaged in self-assessment • Prefer to be paired with a student with the higher ability • Doubtful with the feedback given by underachieving peers • Prefer teacher feedback

Sharing learning outcomes

Students from different achievement levels valued the explication of learning outcomes as it gave them more explicit directions on what to achieve. The teacher explained her rationale for making the outcomes explicit:

> I always think how my students would engage in learning ... I write learning outcomes that are easily understood and break down those complex ones into smaller chunks of outcomes (TI).

This view was shared by her students, and they were all in agreement that the learning outcomes were clear and explicit with "no difficult terms, easy to understand, and they are the starting point for organising our thoughts on how write our essay" (UA1). They agreed that communicating learning outcomes is an indispensable part of the learning process. However, each student group had different perceptions of the approach used by their teacher. For high achieving students, a straightforward discussion of the learning outcomes was enough. This is evident from the response:

> The printed copy is enough…no need to explain. [The learning outcomes] are easy to understand. The more difficult learning outcomes, I think, are the ones to be explained, but after we understand it, she should not repeat it (HA3).

This view of high achievers contrasted with the views of the under-achievers who appreciated the daily discussion of the same learning outcomes for the whole unit of work. They considered it as a critical process for their success: "[H]er constant reminder of the learning outcomes helps me focus my effort. I check those outcomes that I have achieved" (UA2). Another underachieving student commented that the "[L]earning outcomes are like a checklist for me…when my teacher explains them, I am guided which one I have not achieved yet" (UA3). For average students, they stated that "[T]here are learning outcomes that need to be emphasised especially those that require a lot of (uhhmn) thinking…like building a strong argument" (AA2). But other than that, "[S]ome learning outcomes are very straight forward" (AA1)

Although high achieving students found repeating the discussion on learning outcomes not useful for them, and they preferred "to do the class activities right away" (HA5), one student perceived that his teacher was [sharing the learning outcomes] for other students: "I think she is doing it for those who have difficulty in the class. Some of my classmates have difficulty understanding especially the complicated outcomes. That's for them, maybe, but not for me…boring but the others benefit from it" (HA2). He continued by offering a suggestion:

> [M]aybe she needs to separate our classmates who have trouble understanding them and focus with them…if she could do that, I can proceed with my activities while others are being helped to understand the learning outcomes (HA2).

Using exemplars

The students saw the broader strategy of exemplars as very useful "to see [their] teacher's expectations and [they] could use them to model [their] work" (UA3). None of

them cited any negative effects of using exemplars. In fact, when asked, if they preferred to have exemplars in other tasks, they all agreed that other teachers need to use exemplars, "especially for those difficult tasks or those that have higher percentage in our final marks" (HA3).

However, it was evident from the data that the type and use of these exemplars had different effects for the three groups of students. In the literature, the recommended practice is to have a range of exemplars from low to high quality output and the teacher embodied this in her statement:

> …by providing students with a range of exemplars, they would see the common mistakes committed by students…and where the students are mostly penalised. I require all my students to analyse all the exemplars to avoid those mistakes and they can develop better strategies to complete the task (TI).

The findings of this study suggest that this teacher's view of the benefits of analysing a range of exemplars was not shared by many students. The high achieving students wanted to see only the high-quality output as their benchmark because, "[M]y aim is to make sure that I can produce much better output than the best example shown. I focus on that example and identify some areas that I could further work on" (HA4). The high performing students found other exemplars not as useful. This is captured in the statements of two students: "I find it a waste of time analyzing those bad examples" (HA3) because "the bad ones…don't give me any idea to complete the task" (HA2). Also, one student with average ability said, "I know my capacity, and I know the level of performance that I need to (uhhmn) achieve to get a passing mark" (AA1). This was confirmed by another student with average ability in his statement, "I couldn't be bothered looking at the highest levels in the rubrics, because I know I am far from those" (AA3).

The view of high achieving students that the quality of exemplar should match their need was shared also by underachieving students. Giving them exemplars that are higher than what they could possibly achieve is overwhelming. One student in this group said, "when I looked at the best example, I thought, no way, I cannot do that. It scared me, really" (UA1). This means that the match between student perception and the quality of exemplars is an important consideration. As one student summarized:

> When our teacher showed the best example, I thought that's how my output should be as well…I cannot produce a paper like that. When she showed other examples, which are not really good, but still received passing grades, it was a bit of a relief. I analysed those examples and saw some aspects that I could incorporate in my essay… I ended up having a better mark than I thought (UA3).

Eliciting and giving feedback

There was a consensus view among students that they needed feedback to identify areas in their work that required revision. However, students were critical about the content and timing of the feedback and the number of suggestions given.

The feedback they preferred contains "suggestions on how to revise [their] work" (UA1), and they wanted "to read also praises to give [them] an assurance that [they] are on the right track" (HA1). Also, they preferred to receive feedback before the final submission of the output to "have the chance to see if [they] are meeting her expectations or if there are other things that [they] need to do to achieve a higher mark" (UA2).

In terms of the number of key points in the feedback, each group of students had a different preference. For underachieving students, they preferred only a few suggestions as shown in the quote, "[F]eedback always makes me self-conscious especially when there are so many suggestions…three or four suggestions would be manageable" (UA2). The more comments the teacher gave to these students, the more they developed a negative view of feedback. One student pointed out that "the writing in my paper is overwhelming…I felt I was a horrible student. I don't know if she hates me or what" (UA3). However, this view was not shared by high performing students. They preferred to have as many written comments as possible as evident in the statement of student below:

> I want to see all the aspects that I need to improve. My goal is to meet the highest mark and the only way for me to know that is through receiving feedback from my teacher. I mean, I know I am meeting those standards, but still my teacher decides…the more feedback, the merrier (HA2).

This view is also supported by an average performing student:

> I love how my how my teacher identified all the areas that I need to improve. Although it was a bit disappointing to see all the scribbles, but when I reflected on them and made all the changes… my essay (uhmm) turned out to be really good (AA1).

Using self and peer assessment

The teacher employed self-assessment "to engage students in a reflective practice (so they could) see their strengths and more importantly their weaknesses and they can improve their work" (TI). However, the underachieving students preferred to have an expert comment on their paper as UA1 student said, "How will I assess myself? I need an expert telling me what I need to improve my essay." The expert can either be their teacher or classmates that they regard as someone who is more knowledgeable as UA1 student continued:

> My teacher was crazy (laughs), she gave us the rubrics and asked us to assess our essay. I was like, how will I do it? I want someone who can assess my work and recommend something to improve my work. If my teacher can't do it, then maybe one who is better than me (UA1).

Their confidence in the ability of their "better" classmate to give them useful feedback is evident in another student's view: "[A]lthough Peter (not real name) was so mean, he laughed at my writing, but he identified my mistakes and told me what to do to revise my arguments. I did not see those mistakes when I did self-assessment" (UA3).

For high achieving students, "I prefer my teacher gives me feedback than my classmates. I know she will tell me if I am meeting her expectations" (HA1). In peer assessment, they would most likely look at the ability of their peer who will give them feedback. As one student explained:

> I was not happy with my [partner in peer-assessment]. She was telling me that I need to repeat my key ideas in my paper... She did not understand that my summary is the extrapolation of my key arguments. She wants me to use the same words that I used previously. Since then, I ask my teacher to be with someone who can give me better insights (HA5).

This view was shared by an average-performing student as well:

> The effectiveness of peer assessment depends on your role and your pair. I was once with my classmate who was very critical, and he gave me really good suggestions. When it was my turn to give him feedback, I was a bit embarrassed because all I could say was, yeah, great work…and he turned to me and said, is that all? (AA3).

In summary, the responses of students to the four assessment activities used by their teacher were not often aligned to her intent of using sharing learning outcomes, using exemplars, eliciting and giving feedback, and engaging in self and peer assessment to support students in their learning. From the analysis, there was a mismatch between what the teacher expected students to gain, compared with how students perceived and responded to her assessment practices.

DISCUSSION

Our study provides empirical evidence of the association between student achievement levels and their responses to achievement, and how these responses impact on their learning. This association was previously implied only in the literature (e.g., Boud et al., 2015; Colvin et al., 2016; Gynnild, 2003) with no explicit research evidence. Our findings provide three important contributions to understanding how students respond to assessment activities and how these responses influence their future engagement in learning. These findings are interpreted using the biological metaphors of irritability and stimulus-response compatibility. These metaphors serve to assist in the understanding of the implications of the variance in student responses to the same assessment activities.

First, the findings prove that students from different achievement levels respond differently to assessment activities. Students' different responses could explain the variability observed in the extent of the impact of any assessment on student performance (Joughin, 2009). Whilst assessment activities informed by the conceptual framework of AfL generally have positive effects on student outcomes (Baird et al., 2017), we have added evidence that for every assessment activity, there is a particular group of students that accrues more benefit. This highlights an important consideration for differentiating assessment design to ensure that individual students benefit from any assessment implemented (Timperley et al., 2008). This might explain why sometimes

despite strong theoretical and empirical support for the effectiveness of A*f*L, it does not always translate into actual student achievement (Colvin et al., 2016).

Second, our findings clearly established that a teacher's intention in using an assessment strategy is not always realised. This disjunction could explain why assessment practices with robust evidence-base of supporting student learning may not necessarily draw the expected responses from students of a particular achievement level (Boud et al., 2015; Colvin et al., 2016). For example, the use of a range of exemplars is highly recommended in the literature (e.g., Hendry & Anderson, 2013; To & Liu, 2018)) but what is revealed from the evidence in this study is that the effectiveness of using exemplars depends on how students perceive them. The same is true with the use of feedback, which is considered to be the centrepiece of effective learning, but the findings of our study support the earlier observation of Orsmond and Merry (2011) that the teacher's intent of giving feedback is not always accurately perceived by students. Thus, making it ineffective in improving student learning. This supports the notion that a particular group of students selectively engage in assessment to meet their learning and personal agendas (Colvin et al., 2016; Gao, 2009).

Third, this study adds evidence to the research that suggests students' perceptions of the effectiveness of their teacher's assessment practices influence their future motivation toward learning (Cowie, 2009; Lew et al., 2010; Nejad & Mahfoodh, 2019). We have demonstrated that students' achievement levels correlate with how they respond to any assessment activities. The match between the stimulus they received (assessment activities) to their achievement level appeared to either promote or compromise their future engagement. This finding shed light on the issue of student preferences in assessment (Healy et al., 2014). There are no homogenous preferences across all students, and hence, there is no guarantee that even the most credible assessment activity will result in the same outcomes for all students. This finding offers a new lens to understand what factors are critical for differentiation. It is shown that the actual effects of assessment activities on the improvement of student learning is mediated by the variance of students' responses. Close monitoring of these responses provides an opportunity for the teacher to adapt their learning, teaching, and assessment activities to optimise students' learning. This process is part of a teacher's adaptive disposition, which has been linked in the literature to enhanced student learning (Hattie, 2003; Wetzel et al., 2015).

These three contributions of our study have strong implications to the theorisation of effective assessment practices. The significance of these findings can be expressed through the biological metaphor of irritability, which is the characteristics of organisms in being aware of and respond to a stimulus. The stimulus-response compatibility demonstrated in this study mirrors living organisms' ability to sense and respond to any external and internal stimuli in their environment (Haug & Ødegaard, 2015). In principle, favourable stimuli draw positive reactions from the organisms, enabling them to adjust their internal system to achieve physiological balance (Miles & Proctor, 2009). If the stimulus is negligible, organisms tend to ignore it. However, organisms suffer adverse effects if it is unmanageable beyond what their system can successfully adjust.

Our study demonstrated that the right match of assessment activities to students' perceived view of effective assessment, which varies depending on their achievement levels, is critical to eliciting positive responses, impacting their engagement and learning. Thus, this contributes to the theorisation of assessment practices that their effectiveness depends on how teachers design and implement them to draw positive responses from individual students. Adhering to the design of effective assessment practices like sharing learning outcomes, using exemplars, eliciting and giving feedback, and engaging in self and peer assessment, is not enough pre-requisite for optimising the impact of assessment on student learning. The match between teacher intent and students' individual responses is critical, and since students from different achievement levels respond differently, teacher intent is not always translated into actual impact on student learning. Hence, this finding supports the use of differentiated assessment.

Furthermore, the contributions of our study have practical implications to teachers' practices. This implication can be better explained by using the same notion of stimulus-response compatibility. The inherent characteristics of organisms, in this context, the students, is a critical input for teachers to design and implement any assessment strategy. In application of this basic tenet of stimulus-response theory in student learning, teachers need to have a critical view that every assessment stimulus received by a student will draw either a positive or a negative response. To optimise student's learning, the stimulus should be related to their perception of effective assessment practices for the required action to be activated. Otherwise, the response might differ from the expected one (Hommel, 2009; Kornblum et al., 1990). This is clearly shown in our results where the teacher's intent in assessment is not always translated into actual student responses. In fact, it is always mediated by student achievement levels. Some of the earlier evidence has been documented by Lau and Ho (2016), who argue that teacher approaches to helping students either positively or negatively impact upon their achievement depending on how students perceive them.

The principles of irritability and stimulus-response compatibility are not new to the field of assessment. In fact, these concepts inform computer adaptive testing where a preliminary stimulus is provided and the response of the student on the item is used by the computer to adjust the next item, or set of items, to adapt the test to the level of the student (Segall, 2010). Teachers are not computers, but they can adopt an adaptive disposition (Loughland & Alonzo, 2019) that monitors student responses to their initial assessment activity, and then use these responses to develop differentiated assessment activities for students at different achievement levels. This brings the focus of teacher assessment practices to the monitoring of student responses to enhance their learning. In so doing, they will acknowledge that their students' responses are not always aligned with their intentions when employing assessment activities.

CONCLUSION AND RECOMMENDATION

In line with the findings of our study, we offer some recommendations related to practice and future research. For English teachers to optimise the impact of their assessment practices, they have to reflect how each assessment practices can be perceived by students from different achievement levels. Whilst engaging students in

assessment activities, teachers have to monitor how individual students respond to the activity, making sure that their responses positively impact their learning and subsequent engagement in assessment. As shown, a differentiated approach to the actual assessment process is critical for teacher intent to be translated for all students. However, a differentiated approach to assessment based on students' ability level must be done with utmost caution that students will not know about the labels used. The labelling must only be for teachers' input for differentiation and should not be discussed in the class. One potential risk of this labelling is for underachieving students to feel stigmatised and may cause their disengagement. One potential approach to mitigate this drawback to differentiated assessment is to ensure that students have clearly understood the purpose of assessment design and implementation and the critical role of their engagement and response in optimising the impact of assessment on their learning.

In addition, due to the limitation on the number of our sample size, studies with larger sample size are needed to confirm the results of our study. Also, studies using other key learning areas like Science and Mathematics would shed light if the association between student achievement levels and their responses to assessment is content- and context-dependent or universal across key learning areas. Also, quantitative studies are needed to quantify the association between student achievement levels and their responses to assessment strategies. These quantitative studies, either correlational or regression analyses, will establish the extent of association. Furthermore, exploring factors that mediate this association will highlight how to further teacher assessment practices.

Overall, although this study had a limited sample size, and thus the results were not generalizable and warrant further investigation, the findings provide insights that expand the repertoire of student-centred assessment through the adoption of adaptive assessment practices based on student achievement levels. The study indicates how student achievement levels might be used in teacher programs to deliver a student-centred approach to learning and assessment.

REFERENCES

Allen, M. H., Matthews, C. E., & Parsons, S. A. (2013). A second-grade teacher's adaptive teaching during an integrated science-literacy unit. *Teaching and Teacher Education, 35*, 114-125. https://doi.org/https://doi.org/10.1016/j.tate.2013.06.002

Alonzo, D. (2016). *Development and application of a teacher assessment for learning (AfL) literacy tool.* University of New South Wales]. Sydney. http://unsworks.unsw.edu.au/fapi/datastream/unsworks:38345/SOURCE02?view=true

Alonzo, D. (2020). Teacher education and professional development in industry 4.0. The case for building a strong assessment literacy. In Ashadi, J. Priyana, Basikin, A. Triastuti, & N. Putor (Eds.), *Teacher Education and Professional Development in Industry 4.0. 4th International Conference on Teacher Education and Professional Development (InCoTEPD 2019)* (pp. 3-10). Taylor & Francis Group.

Alonzo, D., Labad, V., Bejano, J., & Guerra, F. (2021). The policy-driven dimensions of teacher beliefs about assessment. *Australian Journal of Teacher Education, 46*(3). https://ro.ecu.edu.au/cgi/viewcontent.cgi?article=4761&context=ajte

Baird, J.-A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: Fields apart? *Assessment in Education: Principles, Policy & Practice, 24*(3), 317-350. https://doi.org/10.1080/0969594X.2017.1319337

Barwell, G., & Walker, R. (2009). Peer assessment of oral presentations using clickers: The student experiece. The Student Experience, 32nd HERDSA Annual Conference, Darwin, Australia.

Billany, R., & Billany, T. (2009). A wikied student experience: A collaborative project. The Student Experience, 32nd HERDSA Annual Conference, Darwin, Australia.

Black, P., McCormick, P., James, M., & Pedder, D. (2006). Learning how to learn and assessment for learning. *Research Papers in Education, 2*, 119-132.

Boud, D., Lawson, R., & Thompson, D. G. (2015). The calibration of student judgement through self-assessment: Disruptive effects of assessment patterns. *Higher Education Research & Development, 34*(1), 45-59. https://doi.org/10.1080/07294360.2014.934328

Brown, G. T. L., McInerney, D. M., & Liem, G. A. D. (2009). Student perspectives of assessment. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.), *Student perspective on assessment: What students can tell us about assessment for learning?* (pp. 1-21). Information Age Publishing, Inc.

Bryman, A. (2016). *Social research methods* (4th ed.). Oxford University Press.

Clark, I. (2014). Equitable learning outcomes: Supporting economically and culturally disadvantaged students in 'formative learning environments'. *Improving Schools, 17*(1), 116-126. https://doi.org/10.1177/1365480213519182

Colvin, E., Bacchus, R., Knight, E., & Ritter, L. (2016, 4-7 July 2016). Exploring the way students use rubrics in the context of criterion referenced assessment. Research and Development in Higher Education: The Shape of Higher Education, Fremantle, Australia.

Cowie, B. (2009). My teacher and my friends helped me learn. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.), *Student Perspective on Assessment: What Students can Tell us about Assessment for Learning* (pp. 85-105). Information Age Publishing.

Fives, H., & Barnes, N. (2020). Navigating the complex cognitive task of classroom assessment. *Teaching and Teacher Education, 92*, 103063. https://doi.org/https://doi.org/10.1016/j.tate.2020.103063

Gao, M. (2009). Students' voices in school-based assessment of Hong Kong: A case study. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.), *Student Perspectives on Assessesment* (pp. 107-130). Information Age Publsihing House Inc. .

Ginja, T. G., & Chen, X. (2020). Teacher educators' perspectives and experiences towards differentiated instruction. *International Journal of Instruction, 13*(4), 781-798. https://doi.org/10.29333/iji.2020.13448a

Greve, C., Morris, W., & Huot, B. (2018). Use and misuse of writing rubrics. In *The TESOL Encyclopedia of English Language Teaching* (pp. 1-8). https://doi.org/10.1002/9781118784235.eelt0545

Guo, W., & Wei, J. (2019). Teacher feedback and students' self-regulated learning in Mathematics: A study of chinese secondary students [journal article]. *The Asia-Pacific Education Researcher*. https://doi.org/10.1007/s40299-019-00436-6

Gynnild, V. (2003). On student performance and needs in learning situation: An empirical study. Proceedings of the Learning for an unknown future, 26th HERDSA Annual Conference, Christchurch, New Zealand.

Handley, K., den Outer, B., & Price, M. (2013). Learning to mark: Exemplars, dialogue and participation in assessment communities. *Higher Education Research & Development, 32*(6), 888-900. https://doi.org/10.1080/07294360.2013.806438

Hattie, J. (2003). *Teachers make a difference: What is the research evidence?* Australian Council for Educational Research Annual Conference on Building Teacher Quality, Melbourne.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Haug, B. S., & Ødegaard, M. (2015). Formative assessment and teachers' sensitivity to student responses. *International Journal of Science Education, 37*(4), 629-654. https://doi.org/10.1080/09500693.2014.1003262

Healy, M., McCutcheon, M., & Doran, J. (2014). Student views on assessment activities: Perspectives from their experience on an undergraduate programme. *Accounting Education, 23*(5), 467-482. https://doi.org/10.1080/09639284.2014.949802

Hendry, G. D., & Anderson, J. (2013). Helping students understand the standards of work expected in an essay: using exemplars in mathematics pre-service education classes. *Assessment & Evaluation in Higher Education, 38*(6), 754-768. https://doi.org/10.1080/02602938.2012.703998

Hommel, B. (2009). Action control according to TEC (theory of event coding). *Psychological Research PRPF, 73*(4), 512-526. https://doi.org/10.1007/s00426-009-0234-2

Joughin, G. (2009). Students' expernce of assessment: What don't we know? The Student Experience, 32nd HERDSA Annual Conference, Darwin, Australia.

Kilgour, O. F. G. (1987). Irritability. In O. F. G. Kilgour (Ed.), *Mastering Biology* (pp. 270-303). Macmillan Education UK. https://doi.org/10.1007/978-1-349-09692-3_13

Kite, J., & Phongsavan, P. (2017). Evaluating standards-based assessment rubrics in a postgraduate public health subject. *Assessment & Evaluation in Higher Education, 42*(6), 837-849. https://doi.org/10.1080/02602938.2016.1199773

Kornblum, S., JHasbroueq, & Osman, A. (1990). Dimensional overalp: Cognitive basis for stimulus-response compatability - A model of taxonomy. *Psychological Review, 97*(2), 253-270.

Lau, K.-l., & Ho, E. S.-c. (2016). Reading performance and self-regulated learning of Hong Kong students: What we learnt from PISA 2009 [journal article]. *The Asia-Pacific Education Researcher, 25*(1), 159-171. https://doi.org/10.1007/s40299-015-0246-1

Lew, M. D. N., Alwis, W. A. M., & Schmidt, H. G. (2010). Accuracy of students' self-assessment and their beliefs about its utility. *Assessment & Evaluation in Higher Education, 35*(2), 135-156. https://doi.org/10.1080/02602930802687737

Loughland, T., & Alonzo, D. (2019). Teacher adaptive practices: A key factor in teachers' implementation of assessment for learning. *Australian Journal of Teacher Education, 44*, 18-30.

Magableh, I. S. I., & Abdullah, A. (2020). On the effectiveness of differentiated instruction in the enhancement of Jordanian students' overall achievement. *International Journal of Instruction, 13*(2), 533-548. https://doi.org/10.29333/iji.2020.13237a

Miles, J. D., & Proctor, R. W. (2009). Reducing and restoring stimulus-response compatibility effects by decreasing the discriminability of location words. *Acta Psychologica, 130*(1), 95-102. https://doi.org/10.1016/j.actpsy.2008.10.005

Nejad, A. M., & Mahfoodh, O. H. A. (2019). Assessment of oral presentations: Effectiveness of self-, peer-, and teacher assessments. *International Journal of Instruction, 12*(3), 615-632. https://doi.org/10.29333/iji.2019.12337a

Neuman, W. L. (2013). *Social science methods* (7th ed.). Pearson.

Orsmond, P., & Merry, S. (2009, August 25–29). *Processing tutor feedback: A consideration of qualitative differences in learning outcomes for high and non high achieving students* Fostering Communities of Learners, 13th, EARLI conference, Amsterdam.

Orsmond, P., & Merry, S. (2011). Feedback alignment: effective and ineffective links between tutors' and students' understanding of coursework feedback. *Assessment & Evaluation in Higher Education, 36*(2), 125-136. https://doi.org/10.1080/02602930903201651

Remesal, A. (2009). Accessing primary pupils' conceptions of daily classroom assessment practices. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.), *Student perspectives on asessment: What students can tell us about assessment for learning* (pp. 25-51). Information Age Publishing Inc.

Richez, A., Olivier, G., & Coello, Y. (2016). Stimulus-response compatibility effect in the near-far dimension: A developmental study [Original Research]. *Frontiers in Psychology, 7*(1169). https://doi.org/10.3389/fpsyg.2016.01169

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

Segall, D. O. (2010). Principles of multidimensional adaptive testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of Adaptive Testing* (pp. 57-75). Springer New York. https://doi.org/10.1007/978-0-387-85461-8_3

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189. https://doi.org/10.3102/0034654307313795

Skovholt, K. (2018). Anatomy of a teacher–student feedback encounter. *Teaching and Teacher Education, 69*, 142-153. https://doi.org/https://doi.org/10.1016/j.tate.2017.09.012

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2008). *Teacher professional learning and development: Best evidence synthesis on professional learning and development*. https://researchspace.auckland.ac.nz/bitstream/handle/2292/12537/TPLandDBESentire. pdf?sequence=4

To, J., & Liu, Y. (2018). Using peer and teacher-student exemplar dialogues to unpack assessment standards: challenges and possibilities. *Assessment & Evaluation in Higher Education, 43*(3), 449-460. https://doi.org/10.1080/02602938.2017.1356907

Torres-Guijarro, S., & Bengoechea, M. (2017). Gender differential in self-assessment: A fact neglected in higher education peer and self-assessment techniques. *Higher Education Research & Development, 36*(5), 1072-1084. https://doi.org/10.1080/07294360.2016.1264372

Varsavsky, C., & Rayner, G. (2013). Strategies that challenge: exploring the use of differentiated assessment to challenge high-achieving students in large enrolment undergraduate cohorts. *Assessment & Evaluation in Higher Education, 38*(7), 789-802. https://doi.org/10.1080/02602938.2012.714739

Walker, M. (2015). The quality of written peer feedback on undergraduates' draft answers to an assignment, and the use made of the feedback. *Assessment & Evaluation in Higher Education, 40*(2), 232-247. https://doi.org/10.1080/02602938.2014.898737

Wetzel, A. P., De Arment, S. T., & Reed, E. (2015). Building teacher candidates' adaptive expertise: engaging experienced teachers in prompting reflection. *Reflective Practice, 16*(4), 546-558. https://doi.org/10.1080/14623943.2015.1064380

Wiliam, D. (2017). Assessment and learning: some reflections. *Assessment in Education: Principles, Policy & Practice, 24*(3), 394-403. https://doi.org/10.1080/0969594X.2017.1318108