Assessing written communication skills using a Continua Model of a Guide to Making Judgments (GTMJ)

Peter R. Grainger
University of the Sunshine Coast, peter.grainger@usc.edu.au

Michael Christie
University of the Sunshine Coast

Michael Carey
University of the Sunshine Coast

Assessing written communication skills using a Continua Model of a Guide to Making Judgments (GTMJ)

Abstract

Written communication skills are one of the most assessed criteria in higher education contexts, especially in humanities disciplines, including teacher education. There is a need to research and develop an assessment grading tool (i.e. criteria sheet or rubric) that would assist students in pre-service teacher education programs to better understand and practice written communication and to assist markers when grading academic essays that include this criterion. When rubrics are used the criterion that covers the written communication skills part of the task is often too general to truly assist students to know what they must do in order to obtain the grade to which they aspire. Using substantive conversation in focus group discussions, we defined written communication and designed a criteria sheet using a model known as the Continua model of a Guide to Making Judgments (GTMJ). We found that this tool not only had the potential to help students to better understand key features of good written communication in an academic context, but also to assist assessors to focus on standards descriptors or 'threshold' qualities of written communication when grading students' academic essays.

Introduction

This paper focuses on a particular academic field (teacher education) and the most effective way to assess the kind of written communication that is central to success in higher education, particularly in the humanities subjects. Written assessment, in the form of essays, project reports and theses, is the most common method of determining whether a student has the knowledge and skills needed to graduate. In the sciences and certain professions such as medicine, nursing and teaching, the practicum and field work are essential to complete the student's skill set. But even so, the quality of a student's written communication can affect the grades they achieve. A previous article in this journal has indicated that communicative English proficiency, in both spoken and written modes, is not always of an acceptable standard even at the doctorate level (Li & Vandermensbrugghe 2011). The pressure on academics to improve their students' written communication skills, and thus the quality of higher education, has been increasing since the mid-1990s, as a meta review of the literature on this topic indicates (Harvey & Williams 2010).

This is particularly the case in countries like Australia, the UK and the USA. In Australia, university standards of teaching and learning are overseen by the Tertiary Education Quality and Standards Agency (TEQSA), an independent national regulator of the higher-education sector. TEQSA focuses mainly on the general qualities that a university must have to become a registered provider of higher education, but in the new Higher Education Standards (HES) framework, which was introduced on 1 January 2017, all providers must ensure that they only admit students who have the academic preparation and proficiency in English needed to participate in their intended studies (Australian Government, Tertiary Education Quality and Standards Agency 2015). The HES framework also requires that university lecturers help students to develop and apply the specific and generic knowledge and skills required to graduate in their chosen area and be ready to start employment in that field. For graduates in our focus area (teacher education), this is particularly salient.

Teacher-education program providers within tertiary institutions in a number of countries are required to prepare pre-service teachers to graduate having reached a defined set of professional standards that are set by their state or national accreditation bodies. In England, up until 2012, the General Teaching Council for England (GTCE) acted as the professional body for teachers. Its aim was to improve standards of teaching and learning and regulate professional conduct among teachers. The GTCE was abolished in 2012 and replaced by an executive agency of the Department for Education called the National College for Teaching and Leadership (NCTL), which today funds and monitors initial teacher training that includes 576 Teaching School Alliances. The NCTL also monitors teaching standards and codes of conduct. In the United States the responsibility for teaching registration, standards and codes of conduct devolve to the individual states (National College for Teaching and Leadership 2017).

In Australia the situation was similar to the US until the creation of the Australian Institute for Teaching and School Leadership (AITSL) and the introduction of a national teaching standards agenda. Overall, English-speaking countries require a high level of both verbal and written language proficiency, and the teacher's knowledge and skill in this area needs to be applied across the whole curriculum, not just in the English language and literacy learning area. For this reason, being able to effectively assess written communication skills in initial teacher education, using criterion-referenced assessment sheets or rubrics, is an important factor in ensuring the quality of teaching and learning in this area.

Its importance is exemplified by AITSL's introduction of a set of seven standards that all teachers must understand and apply through assessment tasks. This requirement is incumbent on both pre- and in-service teachers at the graduate, proficient, highly accomplished and lead levels. Standard 2.5 requires teachers at the graduate level to "Know and understand literacy and numeracy teaching strategies and their application in teaching areas" (AITSL 2011). Those who teach the student teachers must also know, understand and apply the AITSL standards if they are to help graduate students meet those standards.

Knowledge of literacy strategies needs to be accompanied by teachers' strong personal literacy (Carey, Christie & Grainger 2015). The government-funded, Louden et al. review (2005) investigated the needs of beginning teachers to teach literacy in schools. The review details the perceptions of beginning teachers and their senior staff colleagues, with senior staff identifying a need for stronger preparation for new graduates in personal literacy, and both cohorts desiring stronger preparation in specific literacy strategies, and in how to use these strategies in teaching and assessment.

The insistence that school teachers need to be more literate and numerate than they currently are has received considerable attention in the media, and has been featured in national and state reviews of teacher training, school curricula and pedagogy over the past decade (Craven et al. 2014; Freebody 2009; Harper & Rennie 2009; Louden et al. 2005; Masters 2009). In response, the Australian government introduced the Literacy and Numeracy Test for Initial Teacher Education (LANTITE) in 2016 (Australian Council for Educational Research 2017). It requires teacher-education students to be within an estimated top 30% of the population for literacy and numeracy before graduating. Carey, Christie and Grainger (2015) offer a further critique of the LANTITE test construct. The test uses multiple-choice questions for ease of marking, cost reduction and rapid turn-around, since the test will be taken by thousands of pre-service teachers every year. The literacy part of the test requires skill in answering reading-comprehension questions based on a text as well as syntax and spelling questions. It does give an indication of how well a person can read, but does not assess the sort of written communication skills that most pre-service teachers are asked to demonstrate during their university studies. The 2017 LANTITE results revealed a national average of 92% of preservice teachers passing the literacy component and 92.3% passing the numeracy component (Department of Education and Training 2018). While this result was publicised as a positive outcome, the result also suggests that 8% of preservice teachers may have graduated and entered the classroom with questionable personal literacy and numeracy.

Key issues in making judgements about written communication

The purpose of our study was, first, to highlight and address a concern with students' written communication skills and how to assess them; and second, to showcase an alternative method to the traditional style of a criterion sheet that could be used to assess these skills. The issue of written communication is relevant to all higher-education contexts, including online contexts (Moon-Heum & Seongmi 2017). The capacity of secondary-school teachers to support general literacy and to teach discipline-specific literacy skills depends on their personal literacy competence. Diagnostic testing of 203 secondary-teaching undergraduates at one Australian university revealed deficiencies in personal literacy competence that could affect their future teaching effectiveness. The sample of undergraduates was tested in spelling, vocabulary and punctuation. Analysis of the results showed high rates of error on general spelling and vocabulary tasks (Moon 2014). As assessors of students, academics all over the world are consistently engaged with this issue when marking and grading student work. Since most of that work is in

written form, the grading tool used to assess the criteria related to the written communication skills of students is of major importance. Hillege et al. (2014) divide these skills up into different categories: academic writing, communication skills and language proficiency.

For assessors, the issue is not just about establishing criteria for what constitutes proficient communication in higher education, but also determining the value attached to these criteria when assessing student work. The assessment literature includes many studies that have described the variability of markers, despite the use of common assessment rubrics (Grainger 2015). This is not a new phenomenon. More than a decade ago Sadler (2005) identified the dominant approach used in judging the quality of student work, writing that it was grounded in the marker's personal expectations (predispositions or biases that markers have regarding the importance of certain criteria) and how the student has performed in relation to other students (i.e. norm referencing).

Jonsson and Svingby's (2007) meta-review of 75 studies found, however, that lecturers can enhance the reliable scoring of performance assessments by using topic-specific rubrics. Their study supports our argument that the use of exemplars and the training of assessors can improve the understanding and application of rubrics and incidentally improve teaching and learning. In our topic-specific area, written communication skills, there is general agreement as to its importance, but significant problems remain in defining and assessing the quality of such written communication (Grainger, Purnell & Zipf 2008).

Schinske and Tanner (2014) cite a number of researchers (Scneider 2013; Jaschik 2009; Higgins, Hartley & Skelton 2004; Brookhart 2006; Pulfrey, Buchs and Butera 2011) who have critiqued criterion-referenced grading tools because of their possible negative side effects, vagueness in the way they define what is valued in the curriculum (criteria) and the way the quality of the performance (the standard) is described (standards descriptors). Grading of student work in tertiary education by means of a matrix model criterion sheet, whereby the grades are selected along a horizontal matrix for standards descriptors aligned with each assessment criterion, is now common (Figure 1).

Criterion	Standard A	Standard B	Standard C	Standard D	Standard E
Ability to…	Standards descriptor	Standards descriptor	Standards descriptor	Standards descriptor	Standards descriptor
Ability to…	Standards descriptor	Standards descriptor	Standards descriptor	Standards descriptor	Standards descriptor

Figure 1. Example matrix model of a rubric

However, Grainger (2015) identified that there is an alternative to this traditional model – the Continua model of a Guide to Making Judgements (GTMJ) (Figure 2) – that has a number of advantages.

The Continua model of a Guide to Making Judgements in university assessment

We have argued that criterion-referenced assessment sheets or rubrics not only have the potential to help students better understand key features of good written communication in an academic

context, but also to assist assessors to focus on standards descriptors or "threshold" qualities of written communication when grading students' academic essays. The key to the usefulness of this tool, however, is that the rubric itself needs to be of high quality. Unfortunately there is a dearth of literature regarding the quality and types of rubrics that could be used in higher education, despite, as argued in this paper, the growing need in the tertiary education sector to provide effective assessment experiences for students, that include not only the quality of the test or task but the provision of a fair and consistent grading and feedback tool; in other words, the need for a well-designed rubric in which the criteria and standards are clear, comprehensible and aligned with the intended learning outcomes specified in the task or test. Grainger and Weir 2016 report a number of common deficiencies that students identify in the design of rubrics, including lack of clarity, subjective descriptors, misalignment and fuzziness in describing what the expected quality is according to the standards identified in the assessment task.

Earlier research by Grainger, Purnell and Zipf 2008 and Sadler 2010 suggested that some student dissatisfaction might be mitigated by the use of the Continua model of a Guide to Making Judgements (Grainger & Weir 2016). This type of rubric has been used extensively in primary and secondary schools and is recommended in a video on standards elaboration developed by the Queensland Curriculum and Assessment Authority (State of Queensland, Queensland Curriculum and Assessment Authority 2017). The Continua model (Figure 2) aims to provide greater satisfaction to students by describing quality, streamlined moderation processes, and giving an additional mechanism for providing feedback to students. However, despite its implementation in the secondary-schooling education sector, the Continua model has not been adopted widely in tertiary settings, and when it has been used, its construction has rarely received collaborative input from staff and students (Christie et al. 2015). This paper seeks to redress what appears to be a gap in the literature regarding research into the use of this type of rubric in higher education. A search of Scopus articles for the last 10 years using the phrase "The use of GTMJ or Continua Model in Higher Education assessment of learning" gave seven hits, but the only relevant one was the 2016 article by Grainger and Weir.
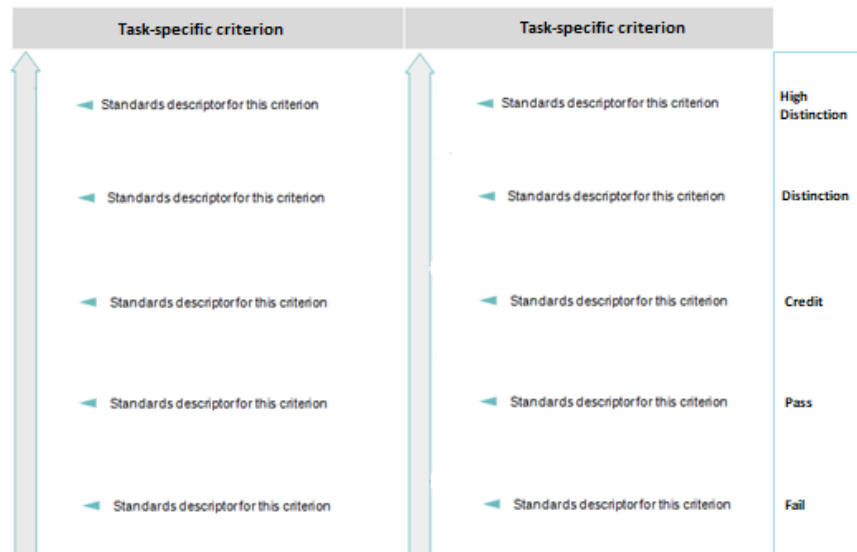

Figure 2: Blank example of the Continua model of a Guide to Making Judgements

The characteristic benefits of the model are summarised by Grainger and Weir (2015) as being: improved design and layout, specifically the use of arrows to depict a continuum of increasing quality rather than discrete measurable units; the aligning of standards descriptors along a continuum, not as a fixed point but as a region of quality; acknowledgement that each criterion does not necessarily have a common number of standards; and finally, the introduction of "nestedness". The nested characteristic ensures that only the defining /discriminating behaviours for each standard are identified, and that the A standard assumes that the student exhibits C standard and B standard behaviours even though they are not identified in the descriptor for each criterion; this provides clear descriptors of performance for students because of the uncluttered and explicit wording of the descriptors, and offers simplified moderation and feedback processes.

Finally, the model emphasises summative, holistic considerations, made across performances in each of the criteria, thus recognising the value of professional judgements to generate a final on-balance judgement about a performance, rather than atomistic, mechanical and additive mechanisms to arrive at a judgement. Grainger, Purnell and Zipf (2008) noted this tendency almost a decade ago:

> …we found that teachers, tutors and lecturers often make judgements based on a holistic assessment of quality, which is then compared with the set criteria and standards and appropriate standards are then awarded in each of the criteria, ensuring the fit is commensurate with the holistic level of achievement awarded (p. 135).

In the area of pre-service teacher education only one study discussed the advantages of using the model as opposed to the traditional matrix style of criteria sheet. That study, conducted in 2016 by Grainger and Weir, was based on interviews with tutors and students. It compared and contrasted the traditional matrix versus the Continua models and reported an analysis of the transcribed data that indicated that the latter is superior to the matrix model due to its positive impact on the three major assessment processes of grading, moderating and feedback. The Continua model's defining characteristics include: the concept of a region of quality represented on a continuum; a focus on defining only new behaviours or the defining quality for each of the standards, thereby avoiding repetition; the concept of nestedness; and design and layout (Grainger & Weir 2016). The major advantage of the model was its focus on only identifying defining or characterising behaviours for each of the standards, something referred to as a *threshold* quality.

The case study – methodology and methods

In our study we researched the efficacy of designing a GTMJ that could be used in the teaching, learning and assessment of the kind of written communication skills we refer to above. Judging by the lack of literature in this area of study, we believe our research is innovative and fills a gap in the literature to address the potential issues with the current models for assessing written communication skills, including vagueness and repetition. We hope that it will add to a collective understanding of how best to assess written communication and provide a more effective grading tool with which to do this. Our research focused on the following:

- Defining criteria for written communication skills;

- Defining the qualities of each of the standards in relation to assessing this written communication;

- Documenting an effective process for writing an assessment rubric;

- Crafting an exemplary rubric for assessing written communication skills in pre-service teacher-education contexts

This paper does not attempt to examine the diverse and contentious field of literacy, or academic literacy specifically. It does, however, provide an exemplar to support academics who are working together to improve their assessment practices.

Although our paper is based in an Australian context, our aim is more general in its application. Its purpose is to determine how a grading tool can measure the quality of written communication, which is an increasingly significant aspect of academic performance in many discipline-specific contexts. Our research methodology could be seen as a type of case study. Yin's (2014, p. 16) definition of a case study in that it is "[a]n empirical inquiry about a contemporary phenomenon" that is set within its real-world context where the boundaries between the phenomenon and the context are not always evident. There are different types of case study, but our research is best defined as an "explanatory case", which, according to Yin, poses "how" or "why" questions about real-life contemporary phenomena.

The phenomenon in focus in this paper is how to assess the quality of the written communication skills of university students consistently and efficaciously. Our case involved collecting the rubrics used by staff in a School of Education in a small regional university in Australia. There are 35 staff in the school and all their rubrics are publicly accessible on our common learning-management system. As noted at the start of this paper, we nevertheless obtained ethical approval to carry out our project. Although the rubrics that we collected (a total of 30) related to different pre-service education courses, all included the assessment of written communication as a common criterion for the tasks that were connected to the rubrics. The authors acted as an expert panel. One of the authors has taught in tertiary-education contexts for over 30 years and written a dozen papers on assessment during that time. As well as this, a third of his career has been spent running pedagogical courses for faculty including workshops on the assessment of learning. The second (an applied linguist) has taught academic writing in tertiary education for 20 years and has published on the craft and assessment of academic writing. The third author is a nationally recognised assessment expert and has published several articles on rubric design. He has been teaching courses on the assessment of learning at both the postgraduate and undergraduate levels for six years. In the first stage of the research each member of the panel analysed the rubrics independently, using a checklist that they had devised during a previous research project (Christie et al. 2015). In the second stage, all the staff of the school were given the opportunity at a school retreat to break into groups and discuss the standards that should be used to help determine the quality of written communication. The authors acted as group leaders and collected the final deliberations of the groups. In the third stage of the research, we, as an expert panel, met three times, with each meeting lasting one hour. These "expert" meetings took the form of substantive conversations that covered the four focus areas of our research (see above), and were audio taped and subsequently transcribed.

Since we hope that our research methods may prove useful for other university staff who wish to undertake research projects to collaboratively craft better rubrics, it is important to briefly define and describe what we mean by substantive conversation. Our aim was to investigate ways to help pre-service teaching students demonstrate their knowledge and skills in written communication in English. To do that, we worked on developing an alternative style of rubric designed to focus attention on how the qualities of written communication skills and English-language proficiency could be measured and graded. Other lecturers could use our methods even if their intention was simply to improve on the existing rubrics that they use in their courses.

We deliberately used the concept of substantive conversations as an organising mechanism for the focus-group discussions that we held amongst ourselves. We define a substantive conversation as one that is characterised by intellectual substance, logical structure and rigorous, respectful and sustained dialogue. The use of this format as a research method is potentially another innovative aspect of our project. We were conscious in our use of it that it could be something that university staff would be able to emulate when they worked together with tutors to improve their course rubrics each year. The term "substantive conversation" most commonly refers to a teaching practice, and has grown in status since Claude Goldenburg (1991) first referred to it as an instructional conversation. Goldenburg argued that true education requires assisting students to understand, appreciate and come to terms with important ideas as they develop a deep understanding for a wide range of issues and questions, and that this can be facilitated by guided instruction (Goldenburg 1991). Although the concept can be traced back to Socrates, it has been relabelled today as substantive conversation and promoted by educationalists and education departments alike. Its popularity can be gauged by the fact that searching for that phrase on Google produces 10, 300, 000 hits within 0.54 seconds. A search of articles related to the concept in Google Scholar produces 686,000 results in 0.43 seconds.

The basis of our substantial conversations was a prior analysis of rubrics that we had produced individually, along with notes from small-group discussions by our peers in an initial education program at a small regional university in Australia. In the three substantial conversations that we held as an expert group, we made use of our data to discuss the definition of written communication in the light of the current literature and our sample rubrics, as well as interrogating the efficacy of the written-communication components of those rubrics.

The purpose of these three substantive conversations was to facilitate the sharing of views among the authors to determine their various definitions and interpretations of criteria, criterion descriptors and holistic judgements about the quality of written communication. We related our views on these issues, and, in particular, how rubrics were used for assessment tasks in our own courses. We concentrated on the relevance of those rubrics in determining the quality of the students' academic communication prior to a task being marked and graded.

We also discussed the differences between the GTMJ and traditional matrix models, and how to best undertake the process of designing the most effective rubric to facilitate the marking, grading, moderating and feedback processes involved in assessing course tasks. We specifically considered what should be identified in that part of a rubric that judged the quality of writing in pre-service education courses. The first author facilitated each session and recorded the substantive conversation that took place. After each meeting, the authors independently analysed the transcripts of what occurred. A number of themes were identified and brought to the subsequent meeting, where they were discussed. The panel agreed on the most relevant data, then extended their analysis. This usually led to a series of new conclusions and decisions.

The practical outcome of our research was the development of a process for creating a rubric with which an assessor could judge written communication and an exemplary model based on that process. The substantial conversations helped the authors fine-tune and expand some of the defining characteristics of a GTMJ model, which are reviewed in the following section.

Threshold qualities and the concept of "nestedness"

The first step in constructing the grading tool was to define the criteria for the task; in this case, there was only one criterion, written communication. The next step was to write standards descriptors for each of the standards we defined; in this case we adopted our official university-wide standards: High Distinction (HD), Distinction (DN), Credit (CR), Pass (P) and Fail (F).

The first standards descriptor to be written was the Pass standard, as this provided a benchmark against which all subsequent descriptors (above and below) can be crafted. We concluded that the Pass standards descriptor must mirror, or at least reflect, the specific criterion that has already been identified in the first step. For example, if the criterion identified is "understands the relationship between x and y", the Pass standards descriptor must capture this. Our reasoning was that if a student cannot meet the requirements of this criterion, this performance is below Pass, and hence, Fail. We refer to this as "threshold quality".

The next step was to craft the standards descriptors for the remaining standards, above and below the Pass level. According to the characteristics of the Continua model, these descriptions should identify new, discriminating and qualitative behaviours that are sufficiently different from each other. Where possible, particularly in reference to the standards above the Fail grade, they should be framed in terms of what students can do, and not what they cannot do, regardless of the standard being described. We adopted the "nested" approach which is characteristic of this model, meaning we did not repeat descriptions of defining behaviours, thereby ensuring that the standards descriptors focused only on the discriminating quality of each standard. This means that the model assumes that the High Distinction student has achieved the behaviours identified at the lower standards (Grainger 2015).

Our discussion around nestedness revealed a potential criticism of this aspect of the model. What if a student did not achieve all characteristics of a previous standard? In the case of a High Distinction standard for any specific criterion, the model assumes that the student will have demonstrated all behaviours of all standards leading up to this standard. Our discussions resulted in the conclusion that an assessor could use an on-balance professional judgement, considering the relative importance of the criterion and the student behaviour that was not achieved, to decide if the student was able to achieve the High Distinction standard. This is in accordance with the thoughts of Wyatt-Smith and Klenowski (2013, p.35), who stated, "In the assessment of student work using standards and marking guides, it is generally accepted that human judgement is central to the assessment process."

We noted that a characteristic of the Continua model was the lack of a requirement to identify standards descriptors at every standard, due to the nature of the continuum of quality, rather than a defined point of quality. To identify the descriptors, we scrutinised words used to discriminate behaviours that lacked clarity. For example, the word "very" is often used by assessors to discriminate standards (for example, alongside "insightful"), but the word itself demands definition. How much is "very"? It is more explicit to use adjectives that graduate in intensity for each standards descriptor instead of using adverbial intensifiers such as "very". We shared our ideas of what the visual evidence would look like in student work to come up with a standards descriptor for each standard. This was based on our "tacit knowledge" (Sadler 2005) of what we would expect from students at each standard. This process is summarised in Table 1.

Table 1. Process used to develop a Continua model of a Guide to Making Judgements

Stage	Key activities for this stage
1	Define the criterion through substantive conversation
2	Identify, through substantive conversations, the key aspects of quality that will be evidenced in student work for this criterion (e.g. spelling, grammar, referencing)
3	Write the standards descriptor for these key threshold qualities to define the PASS level
4	Working down the continuum, write the FAIL standards descriptor, framed positively, and describing what a student can do (not what they can't do)
5	Working up the continuum write the standards descriptor for the key threshold qualities that characterise the HIGH DISTINCTION level
6	Write standards descriptors for the remaining standards, recalling the concept of nestedness

Creating a GTMJ tool for higher education

Our analysis of the data, derived from the substantive conversations, enabled us to create a Continua model of a Guide to Making Judgements (GTMJ) for assessing written communication. Our context was a pre-service teacher-education program. We defined written communication as the skills or abilities that a student demonstrates regarding writing in an academic genre. This included mechanical aspects of form such as spelling, sentence construction and grammar, as well as the functional ability to write according to the conventions of the academic genre, including referencing protocols and less significant technical aspects of the criterion, such as paragraphing, word counts and unified font sizes.

We concluded that a Pass standard would show the ability to write according to the defined key academic referencing protocols of the genre (in our context, APA style). In addition, a student at this standard would be able to construct sentences that, despite minor errors, were generally comprehensible. In addition, a student at this standard would show that they had read the cited literature and could refer to this literature appropriately in-text. We recognised that the Pass student makes errors but provides enough evidence of competence to meet the criterion.

In crafting the Fail standard, we referred to the characteristics of the Continua model and attempted to write descriptors that were positively framed; that is, defining what a student can do, rather than what they cannot do, which, in our collective assessment experience in tertiary education of over 50 years, was the norm. Our tacit knowledge described this standard as being typified by unsupported statements and references to unknown authored works, as well as a myriad of language errors that made it difficult for the reader to recover the meaning.

Our next step was to define the High Distinction standard, which we found to be a relatively simple task, as we all agreed that typically this standard is met by high-calibre writing. We teased this out by identifying the characteristics of high-quality writing. We defined these characteristics as cohesion, created with lexical ties (such as repeated words, synonyms, pronouns and

conjunctions) used to develop topics and guide the reader between sentences and paragraphs, and the ability to synthesise sources to enhance the arguments being presented.

Our final two standards (Distinction and Credit) proved difficult to craft. We focused on the discourse macro-structure, specifically the use of introductions and conclusions, as characteristic of the standards. We also agreed that a student who achieved these standards would be more consistently accurate in their referencing, grammar and spelling, as well as being more detailed in their coverage of the topic and explanations. A major piece of evidence at these levels was identified as the use of sources, with these students reading more widely and beyond the seminal works often cited by academics. When deciding how best to discriminate between Credit and Distinction, we identified the key evidence as the demonstration of higher-order thinking skills such as the ability to compare and contrast different views.

We recognised that the Continua model only defined new and discriminating behaviours at each standard, and hence we were able to avoid repetition of descriptions, thus giving the model an un-cluttered feel and, more importantly, a focus on the typical behaviour within each standard. The resulting Guide to Making Judgements is presented in Figure 3.
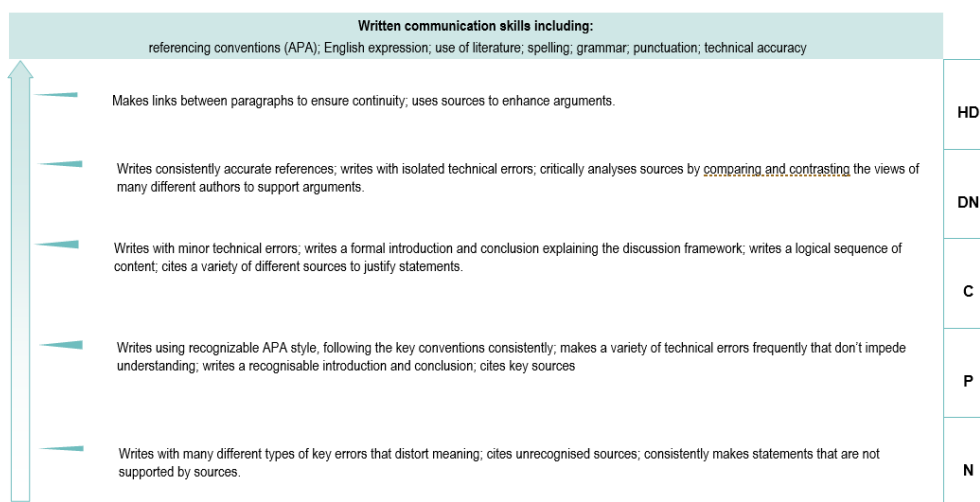

Figure 3. Exemplar of a Continua Model of a Guide to Making Judgements (GTMJ) to assess written communication skills

Findings and discussion

Our conversations identified several issues when constructing this grading tool. Initial discussions centred on defining the criterion, standards and written communication, and on determining what aspects of this criterion would be used to judge achievement. Some assessors wanted to assess different aspects discretely in their own right rather than to identify just one criterion that subsumed different aspects. The assessment literature refers to this as fine-grained versus coarse-grained distinctions (Grainger 2015), and has identified differences between raters in what they focused on in the assessment of writing in terms of organisation, content, grammar and handwriting (Vaughan 1991).

The difficulty of determining criteria, which was the first step in our process, resonates with much of the assessment literature, which has acknowledged this difficulty due to the different interpretations of words leading to fuzzy definitions that represent individual mental constructs (Sadler 1985). Consequently, Sadler (1985) advocated for global or holistic evaluation as the fundamental basis for reflection and the source or starting point for identifying subsequent criteria. We discussed the inability of our tool to explicitly discriminate our different expectations (i.e. standards) of students in relation to their year level. In other words, we acknowledged that our expectations grew as students approached graduation, and yet we were unable to identify this in the standards descriptors. In addition to the difficulty in constructing criteria, assessment literature points to the same difficulty when identifying standards (Wyatt-Smith & Klenowski 2013), which is deeply rooted in evaluative experiences.

We acknowledged that our tool could serve a first-year university course as well as a master's level course. We concluded that this expectation can only be identified via substantive conversations, either prior to or at the time of moderation.

We disagreed on what constituted "quality" at each of the standards for a specific cohort. In other words, even though we agreed on the criterion, we had different views on what each standards descriptor would include as characteristic of the level that we referred to as a "threshold quality". Markers have implicit understandings of quality, based on their own predispositions (Grainger, Purnell and Zipf 2008). In this regard we enjoyed the freedom associated with ticking along a continuum of quality, rather than specific points, but we found ourselves wanting to write descriptors for each of the standards.

We discussed the concept of "nestedness": specifically, the fact that some academics might prefer to repeat descriptions of behaviour right up the continuum, in the fear that this evidence might be "forgotten" when assigning the grade to this criterion. We noted the necessity to explain this aspect of nestedness in detail and consistently to students prior to assessment and to explain other markers involved in assessing the task. There was a desire to include numerical values (i.e. numbers) when describing, for example, the number of references required for each standard, but we eventually decided that the focus should be on quality, not quantity, and in this regard the quality of the author being cited became paramount. Two of the lecturers in the study wanted to include standard type words in the criterion descriptor, especially for this master's level, as there was an expectation that a master's degree student should be able to reference accurately all the time at the Pass standard, having already completed an undergraduate degree. An example of this was the aspect of mastery of APA protocols, which we all agreed was an expectation of the Pass standard for the master's level but not for the undergraduate level.

Upon reflecting on our collective experience, our discussions around the specifics of the Continua model may be attributed to our own predispositions towards traditional models of criteria sheets; specifically, the matrix model. In terms of this, we identify the work of Borg (2004), who describes an apprenticeship of observation whereby pre-service teachers are influenced by and find it difficult to evaluate new experiences. We likened ourselves to pre-service teachers in this regard, having experienced many years of using traditional style matrix models that affect our ability to accept the characteristics of the "new" Continua model. As evidence of this, two of the authors articulated the need to write a descriptor against every standard, noting the probable student resistance to the concept that a standard could be achieved without a specific description of the standard.

An important feature of the rubric that ensued from our discussions was the identification of a "threshold quality" for each of the standards. This feature ensured that the grading tool we developed was uncluttered by multiple, repeated descriptors, and was visually appealing, focused only on discernible differences or defining behaviours; we felt that this would hence ease the subsequent marking, moderation and feedback processes.

Conclusions

Our study filled a gap in the literature by focusing on the assessment of written communication in a pre-service education context using an alternative model to the matrix style of criterion sheet, known as the Continua model of a Guide to Making Judgements. We specified how these written communication skills can be defined and, importantly, how they can be judged using a grading tool that provides clarity and explicit, well-defined standards of achievement. The process we used, which included substantive conversations, is transferable to different contexts other than pre-service teacher education. In going through this process, we identified debates on subjective versus objective marking, grading and criteria versus holistic/connoisseurship assessment approaches and the development of assessment criteria and standards.

Foremost was our focus on assessing written communication in the context of an ever-increasing accountability agenda in higher education, driven by external accreditation bodies and, in our specific context, the regulatory bodies of the education context. In short, we acknowledged that written communication is almost a mandatory criterion for assessment in our context. Hence, for reasons of consistency, we wanted to define these skills and subsequently develop a process for creating a grading tool (Guide to Making Judgements, criterion sheet) that assessed these skills, leading to the development of a product that could be used in education courses initially, but would also be relevant to all higher-education programs and courses that assess written communication. We were conscious of the assessment literature that has been very critical of assessment processes such as grading, moderation and feedback.

We set about first to discuss how to assess written communication, and then participated in a deliberately sequenced process to develop the grading tool. In developing the tool, we used a new model of a criterion sheet, described by Grainger (2015) as the Continua model of a Guide to Making Judgements or GTMJ. In developing our rubric, we noted some key aspects of this model; primarily its focus on only defining key characteristics for each standard, written as standards descriptors. In short, the model forces assessors to define student behaviour or qualities of behaviour that are typically shown at each of the standards. We refer to these as threshold qualities. This focus on qualities affects not just the actual grading process at "point of service", but also the post-grading quality-assurance mechanism of moderation. Focusing ahead of time on what evidence will be seen in student work facilitates a smoother moderation process. This is because the substantive conversation between markers about what constitutes quality has already been had prior to marking. We also noted that this early focus would reduce the need for explicit feedback, as the standards descriptors are themselves so explicit that students can be in little doubt about what evidence contributed to the grade; this in essence creates a feed-forward process.

Our pre- and post-rubric development discussions revealed some interesting reflections about our assessment process. Two in particular stand out. The first is that academics construct grading tools based on their own tacit knowledge of the task, the task requirements and their experiences in grading. Regarding written communication, we noted that we had quite defined, explicit notions of what this criterion meant, and the "qualities" we would expect in student work. Despite this,

participation in the process to create the tool revealed varied expectations of what the standards descriptions should look like, variations that could only be overcome via substantive conversations, which routinely occur after the student work has been marked.

We recommend, therefore, that substantive conversations occur before the Guide to Making Judgements is created and before the student work is marked. This would facilitate better understanding of the tacit knowledge that each assessor brings to the marking process, which would enable better descriptions of quality on the GTMJ. The substantive conversations should involve all those associated with marking the student work, especially sessional or casual staff members. We believe this reflects our belief that the formal process of post-grading assessment moderation was of equivalent importance to the pre-grading substantive conversations we held to craft the criterion sheet used to evaluate student performance.

The second point, involving all future assessors, is an important one given the increasing number of casual employees within the academic workforce. Casual or sessional staff teach, mark and grade anywhere between 50% and 80% of work in Australia (Grainger 2015). Our own collective experiences across seven different universities suggest that the process for creating an assessment rubric does not typically involve input from sessional staff, and is often done by the Course Coordinator in isolation and usually at least six months before implementation. We are not suggesting that these conversations are not taking place in some institutions, and we acknowledge the logistical difficulties and financial implications of including casual staff in the creation of these rubrics.

We acknowledge the need for further widespread testing of the Continua model of a Guide to Making Judgements across different disciplines to ensure its widespread applicability. This study provides further evidence that the construction of a quality criterion sheet is a difficult process and driven by tacit understanding that is not always shared, even by experienced assessors. In contexts of teacher education, in both the pre-service and post-graduation stages, assessing written communication is of growing importance, and the ability to collaboratively design a quality grading tool that can be modelled by assessors for their students' demands further research.

The paper identified a gap in the literature in regard to the use of a Continua model of rubric in higher education and conducted a case study within a small School of Education in Australia. The case study involved the analysis of 30 rubrics for tasks that included written communication as a criterion. By analysing the rubrics and subjecting that analysis to further discussion in a series of three substantive conversations among a panel of experts, a Continua Model of a Guide for Making Judgements was created. This model provides new insights and practical assistance in improving rubrics that include among their criteria the assessment of written communication skills.

References

Australian Government, Tertiary Education Quality and Standards Agency 2015, Higher Education Standards Framework (Threshold Standards) 2015, viewed 15 May 2018, http://www.teqsa.gov.au/teqsa-contextual-overview-hes-framework.

Australian Institute for Teaching and School Leadership 2011, 'Australian Professional Standards for Teachers,' viewed 15 May 2018, http://www.aitsl.edu.au/docs/default-source/apst-resources/australian_professional_standard_for_teachers_final.pdf.

Australian Council for Educational Research 2017, 'Literacy and Numeracy Test for Initial Teacher Education Students,' viewed 15 May 2018, https://teacheredtest.acer.edu.au/.

Borg, M 2004. 'The apprenticeship of observation', *ELT Journal*, vol. 58, no. 3, pp. 274-276.

Brookhart, SM 2008. *How to Give Effective Feedback to Your Students*, Association for Supervision and Curriculum Development, Alexandria, VA.

Carey, MD, Christie, M & Grainger, P 2015, 'What benefits can be derived from teaching knowledge about language to preservice teachers?', *Australian Journal of Teacher Education,* vol. 40, no. 9, pp. 16-30.

Christie, M, Grainger, P, Dahlgren, R, Call, K, Heck, D & Simon, S 2015, 'Improving the quality of assessment grading tools in Master of Education courses: a comparative case study in the Scholarship of Teaching and Learning', *Journal of the Scholarship of Teaching and Learning,* vol. 15, no. 5, pp. 22-35.

Craven, G, Beswick, K, Fleming, J, Fletcher, T, Green, M, Jensen, B, Leinonen, E & Rickards, F 2014, 'Action Now: Classroom Ready Teachers', *Teacher Education Ministerial Advisory Group Final Report*, viewed 15 May 2018, http://www.studentsfirst.gov.au/teacher-education-ministerial-advisory-group.

Department of Education and Training 2018, 'Test ensuring teaching is up to scratch', *Ministers for the Department of Education and Training Media Centre*, viewed 15 May 2018, https://ministers.education.gov.au/birmingham/test-ensuring-teaching-scratch.

Freebody, P 2009, 'Literacy education in school: Research perspectives from the past, for the future', *Australian Education Review*, Victoria, Australian Council for Educational Research, viewed 15 May 2018, http://www.acer.edu.au/documents/AER52.pdf.

Grainger, P, Purnell, K & Zipf, R 2008, 'Assessing the quality of student work: The role of "substantive conversation" between markers', *Assessment and Evaluation in Higher Education,* vol. 33, no. 2, pp.133-142.

Grainger, P 2015, 'How do preservice teacher education students respond to assessment feedback?', *Assessment and Evaluation in Higher Education*, DOI: 10.1080/02602938.2015.1096322

Grainger, P & Weir, K 2016, 'An alternative grading tool for enhancing assessment practise and quality assurance in higher education', *Innovations in Education & Teaching International,* vol. 53, no.1.

Harper, H & Rennie, J 2009, 'I had to go out and get myself a book on grammar: A study of preservice teachers' knowledge about language', *Australian Journal of Language and Literacy,* vol. 32, no. 1, pp. 22-37.

Harvey, L & Williams, J 2010, 'Fifteen Years of Quality in Higher Education (Part Two)', *Quality in Higher Education*, vol. 16, no. 2, pp. 81-113.

Higgins, R, Hartley, P & Skelton, A 2002, 'The conscientious consumer: reconsidering the role of assessment feedback in student learning', *Studies in Higher Education,* vol. 27, pp. 53-64.

Hillege, SP, Catterall, J, Beale, BL & Stewart, L 2014, 'Discipline matters: embedding academic literacies into an undergraduate nursing program', *Nurse Education in Practice,* vol. 14, no. 6, pp. 686-691.

Jaschik, S 2009, *Imagining College without Grades,* viewed 20 February 2014, http//:www.insidehighered.com/news/2009/01/22/grades.

Jonsson, A & Svingby, G 2007, 'The use of scoring rubrics: Reliability, validity and educational consequences', *Educational Research Review,* vol. 2, pp.130-144.

Li, LY & Vandermensbrugghe, J 2011, 'Supporting the Thesis Writing Process of International Research Students through an Ongoing Writing Group', *Innovations in Education and Teaching International*, vol.48, no. 2, pp.195-205.

Louden, W, Rohl, M Gore, J, Greaves, D, Mcintosh, A, Wright, R, Siemon, D & House, H 2005, *Prepared to teach: An investigation into the preparation of teachers to teach literacy and numeracy*, Australian Government Department of Education, Science and Training, Canberra.

Masters, GN 2009, 'A Shared Challenge: Improving Literacy, Numeracy and Science Learning in Queensland Primary Schools', Australian Council for Educational Research.

Moon, B 2014, 'The Literacy Skills of Secondary Teaching Undergraduates: Results of Diagnostic Testing and a Discussion of Findings', *Australian Journal of Teacher Education*, vol. 39, no. 12, pp. 111-130.

Moon-Heum, C & Seongmi, L 2017, 'Using regulation activities to improve undergraduate collaborative writing on wikis', *Innovations in Education and Teaching International*, vol. 54, pp. 53-61.

National College for Teaching and Leadership 2017, viewed 15 May 2018, https://www.gov.uk/government/organisations/national-college-for-teaching-and-leadership.

Pulfrey, C, Buchs, C & Butera, F 2011, 'Why grades engender performance-avoidance goals: the mediating role of autonomous motivation', *Educational Psychology*, vol. 103, no. 3, pp. 683-700.

Sadler, DR 1985, 'The origins and functions of evaluative criteria', *Educational Theory*, vol. 35, no. 3, pp. 285-97.

Sadler, DR 2005, 'Interpretations of criteria-based assessment and grading in higher education', *Assessment & Evaluation in Higher Education*, vol. 30, no. 2, pp. 175-194.

Sadler, DR 2010, 'Fidelity as a precondition for integrity in grading academic achievement' (review), *Assessment & Evaluation in Higher Education*, vol. 35, no. 6, pp.727-43.

Schinske, J & Tanner, K 2014, 'Teaching more by grading less (or differently) in CBE', *Life Science Education,* vol. 13, no. 2, pp.159-166.

State of Queensland, Queensland Curriculum and Assessment Authority 2017, 'Using the Standards Elaborations (videos),' viewed 15 May 2018, https://www.qcaa.qld.edu.au/p-10/aciq/standards-elaborations/using-videos.

Tertiary Education Quality and Standards Agency 2017, viewed 15 May 2018, http://www.teqsa.gov.au/.

Vaughan, C 1991, 'Holistic assessment: What goes on in the rater's mind', *Assessing second language writing in academic contexts,* Ablex Publishing Corporation, Norwood, NJ.

Wyatt-Smith, C & Klenowski, V 2013, 'Explicit, latent and meta-criteria: types of criteria at play in professional judgement practice', *Assessment in Education: Principles, Policy & Practice*, vol. 20, no.1, pp. 35-52.

Yin, RK 2014, *Case study research: Design and methods*, Sage, Los Angeles.