

‘It’s Hard to Define Good Writing, but I Recognise it when I See it’: Can Consensus-Based Assessment Evaluate the Teaching of Writing?

Ursula Canton
Glasgow Caledonian University, UK

Abstract

In a Higher Education environment where evidence-based practice and accountability are highly valued, most writing practitioners will be familiar with direct requests or less tangible pressures to demonstrate that their teaching has a positive impact on students’ writing skills. Although such evaluations are not devoid of risk and the need for them is contested, it can be argued that it is better to engage with them, as this can avoid the danger of overly simplistic forms of measurements being imposed. The current paper engages with this question by proposing the conceptual basis for a new measurement tool. Based on Amabile’s Consensual Assessment Technique (CAT), developed to assess creativity, the tool develops the idea of consensual assessment of writing as a methodology that can provide robust data through systematic measurement. At the same time, I argue consensual assessment reflects the evaluation of writing in real life situations more closely than many of the methodologies for writing assessment used in other contexts, primarily large scale tests. As such, it would allow writing practitioners to go beyond ethnographic methods, or self-reporting, in order to obtain greater insight into the ways in which their teaching helps change students’ actual writing, without sacrificing the complexity of writing as social interaction, which is fundamental to an academic literacies approach.

Introduction: The Value and Challenges of Assessing the Impact of Teaching Writing

In the UK, the insight that ‘the power to write well [...] is an essential tool for survival in a sophisticated world’ (Ahmad and McMahon 2006: 4) has increasingly led to ‘an acknowledgement, albeit often implicit, of the university’s responsibility to develop students’ writing’ (Jones 2004: 254). Yet, the ability to communicate efficiently in writing is not often included explicitly in learning outcomes for specific modules, even if it features prominently in many lists of ‘graduate attributes’ developed by universities or reported in surveys among graduate employers (Diamond et al. 2011: 8). Similarly, writing is often used as a tool for assessment, but the explicit teaching of writing is not always embedded in programme structures. As a result, the teaching of writing in Higher Education (HE) is not part of the formal quality assurance processes applied to programmes or modules.

The advantages and disadvantages brought by this position “on the margin” of formal processes are examined in the first part of this paper. Concluding that evaluating the impact of the teaching of writing is important, the second part then reviews existing measuring tools, and establishes their limitations for assessing the notion of writing as social interaction that underlies teaching practice based on an academic literacies framework. As a solution, the paper develops the conceptual basis for a new approach to assessing writing based on consensus among readers, presented in the third part. Inspired by Amabile’s (1982, 1996) and Hennessey,

Amabile, and Mueller's (2011) work on creativity, consensual assessment offers a radical departure from many previous methods, as it uses an operational definition of writing that focuses less on the qualities of texts, but primarily on the effect texts can have on readers. Through this change in focus, it could facilitate the collection of robust data on writers' performance – and improvement – in varying communicative situations.

Not being part of formal quality assurance processes allows writing practitioners greater freedom in teaching and, to some degree, shelters us from the 'deceptive simplicity' of tightly formulated learning outcomes, giving us the opportunity to approach teaching – and writing – in its full complexity of 'interests and values, policies and practices' (James 2005: 93). Nonetheless, this does not exempt writing centres, learning centres and other institutions responsible for the teaching of writing from any form of accountability towards the public or fee-paying students, who fund them through taxes or tuition fees, and towards their institutions that allocate these resources. This accountability is clearly perceived by writing practitioners, as the written notes elicited from workshop participants at the biennial conference of the European Association for the Teaching of Academic Writing (EATAW) confirm: many participants expressed the need to document their impact, as 'justification of what we do as basis for decision makers,' evidence to 'keep or ask for more resources from an institution' or as a general step towards 'accountability' (EATAW Members 2017).

Accountability and the need to document impact can, however, be a double edged sword. It does not automatically lead to improvements and can have negative side effects (see, for example, Hendry and Dean 2002), which were clearly recognised by some participants of the workshop. There is the danger that assessing the impact of teaching 'may lead to a bias towards teaching that is measurable' or that teaching projects 'with potential if developed' may be abandoned after a negative initial evaluation (EATAW Members 2017). These risks can increase writing practitioners' inclination to evade the external pressures that demand evidence of our impact and to enjoy the relative liberty afforded by our position outside the strict quality assurance procedures for accredited programmes of study. Such a stance may, however, be short sighted: as one EATAW member pointed out, it was good to have evaluation data at hand, 'just in case – decision makers may suddenly ask' (2017). More importantly, such an outright rejection ignores another powerful reason in favour of assessing the impact of teaching: as an overwhelming majority of the workshop participants attest, it offers them an additional source of information on students' understanding and, therefore, helps them improve their own teaching practice. Understanding whether our actions make a difference is a fundamental human concern, but is particularly important for a profession whose *raison d'être* is to enable others to change their level of knowledge, understanding or ability, or whose moral obligation is to 'facilitate learning' according to Regan (2012). Assessing whether our teaching has the desired impact is thus an important professional concern, albeit a rather complex and difficult one.

The Challenge of Assessing Impact

The complexity of assessing the impact of writing teaching results from the many-faceted construct we are teaching: writing can be regarded as a product (text), a process (of writing a text) and a skill (the ability to produce texts). The ability to navigate the process and produce a successful text depends on 'many external factors' (EATAW Members 2017), so improvements are not independent of context either, nor can they be attributed directly to a specific teaching intervention. In addition, changes in writers' skills that lead to changes in their writing process or in the texts they produce can be slow, which means that the temporal aspect of any evaluation needs to be carefully considered. Any evaluation that closely follows a teaching intervention might miss important longer-term changes; yet, the longer the temporal distance between teaching and evaluation, the more likely it becomes that changes are also influenced by other factors. Many of the reservations against assessment expressed by writing practitioners can be attributed to the fear that formal evaluation of the impact of their teaching could disregard the complexity of writing. Results from a simplistic approach to evaluation would not only be meaningless as feedback on teaching practice, but could lead, in the worst case, to

a misrepresentation of the impact of teaching: short-term measurements could lead to a negative evaluation of teaching that could bring long-term benefit (EATAW Members 2017).

In addition to these possible complications, assessing the impact of teaching in such a complex construct is very time-consuming. It seems impossible, under these circumstances, to avoid the question of whether such a potentially risky and resource-intensive endeavour is worth undertaking. The benefits of understanding the impact of teaching interventions identified above, on the other hand, are significant. Accepting the challenge of evaluating whether the teaching of writing has a measurable impact on student writers could improve teaching practice and justify our work towards others as part of a wider culture of accountability. If the challenge is accepted, the first and foremost question then becomes which tools and instruments can be used for this endeavour. After reviewing existing instruments and identifying the problems associated with them, this paper proposes a new approach to evaluating writing and improvements in writing as part of measuring the impact of writing support.

Outline

In this paper I propose the rationale for a new tool to evaluate writing. First I review studies that evaluate the impact of writing classes and support, many of which place strong emphasis on the writer. This emphasis can be considered inappropriate for teaching focused on the interactive nature of writing, as it prioritises the writer's perception of her work at the expense of that of the reader, the second "interactor." Searching for less writer-focused forms of assessment in related disciplines, I then examine whether these measurement tools could capture the desired impact of teaching and supporting writing in HE. While many of these tools offer great reliability, their validity is questionable in our context, as the construct they measure is fundamentally different to the theoretical framework underlying our teaching. I explain this discrepancy between the conceptual basis of our teaching and of large scale writing assessment and develop the theoretical groundwork for an alternative approach, based on Amabile's work in the field of creativity. This relies on the concept of consensual assessment and attempts to include readers and their judgement in a manner that reflects our definition of writing as social interaction.

Current Practice in Assessing Teaching Impact

Most studies that assess the impact of teaching writing in the UK regard the writer as the ultimate expert for her own writing progress, and rely on self-reporting to evaluate changes in students' attitudes towards writing and their confidence in their writing skills. To reduce the limitations of self-reporting, these evaluations are often combined with ethnographic methods, such as observations and reflections from lecturers (Bharutram and McKenna 2006, Li and Vandermensbrugge 2001) or interviews (Sengupta and Leung 2002). This combination of different ethnographic approaches with a strong focus on interviews resembles the methodologies used to develop academic literacies (e.g. Lea and Street 1998, Lillis 2001), which today is the most prominent conceptual basis for much of the writing support in the UK. Both the development and the evaluation of writing support in HE thus focus on the writer. Readers, as the second active party in the communication process, or texts, as the product of the writing process, have rarely been involved in evaluation, despite the fact that the academic literacies approach is based on a definition of literacies as 'cultural and social practice' (Lea and Street 1998: 158; emphasis added). This means that, over a decade after publication, Thonus' observation that writing centre assessment is rarely 'connected with assessments of [...] the quality of students' writing' (2002: 112) still applies, at least in terms of direct measurement.

With regard to assessment, writing differs greatly from other skills and knowledge taught at university. It is unlikely that the impact of teaching mathematics would be measured mainly, or solely by asking students' about their levels of confidence in approaching equations, for example. This focus cannot only be explained by the complexity of the skill students are meant to acquire – performance assessment as an indication of ability is used for other complex skills, including ill-defined ones such as design. An alternative explanation can be found in administrative structures in HE, which often place writing support outside core studies.

Similarly, its assessment mirrors that of other university *services*, in which students can indicate their satisfaction with the support they receive. Asking student writers about their own perception of the support they received and the impact this has had on their writing is the logical extension of this approach. While this perception is certainly an important part of assessing the impact of teaching, relying solely on 'the data from student comments' or other ethnographic methods' is 'not sufficiently incisive' (Leibowitz et al. 1997: 15) and unsatisfactory from a pedagogical point of view, as it does not illuminate whether any potential changes in students' perception of the writing process are reflected in their interaction with readers. Whether teaching improves communication with readers thus remains unclear.

One possibility to address this shortcoming is to use indirect measures of changes in students' writing. For example, Nzekwe-Excel (2014) examines the correlation between attendance at writing workshops and essay marks. The use of circumstantial factors is, however, problematic. Firstly, it does not acknowledge the influence of the many potential confounding variables that influence marks: although markers can be considered the intended readers for a text, the question of whether students communicate successfully with them through their text is only one of the many factors that will influence their marks. Improved marks cannot be directly attributed to improved writing. Secondly, it ignores potential selection bias: it can also be expected that students who attend additional writing sessions are more committed and might, therefore, achieve better marks even without this input. These disadvantages of using circumstantial measures also render many internal reviews problematic, as well as other forms of writing centre assessment that rely on marks or progression and retention figures as indirect measurement of improved writing abilities.

Assessment in Related Disciplines

Large Scale Writing Tests

An alternative approach is assessment of changes in written texts themselves. This approach is widely used in large scale assessments for second language writers, as well as college aptitude tests in the US. Among the tests used for college admissions in the US, the Scholastic Assessment Test (SAT) includes writing a short essay (College Board n.d a) and the American College Test (ACT) offers writing an original text in response to a prompt as an optional component (ACT n.d.). In addition to college admission, the scores from these tests, or from alternative, locally developed tests, can be used to place students in different levels of composition and rhetoric courses. The emphasis on using these large scale tests for administrative purposes (Wind and Engelhard 2013) means, first of all, that they are oriented towards identifying differences in the writing abilities of a student cohort, and not toward tracing changes in the writing of individual students. Writing tests for second language speakers, such as Cambridge Exams, TOEFL or IELTS, present the same problem (Banerjee and Wall 2006).

Since the assessment criteria for these admissions and language tests are remarkably similar, they also share the second problem that makes them unsuitable to measuring the impact of the teaching of writing in HE. Despite including criteria like 'task achievement' (College Board n.d. b) or 'understanding of the task' (IELTS n.d. a), neither SAT, nor IELTS provide much information about the communicative situation in which the tasks are presented (with the exception of the letter writing task in IELTS General Training tests (IELTS n.d. b), which usually includes a short introduction of the situation in which the letter is to be written, e.g. a letter to a friend who has agreed to house- and pet-sit). The more academic texts ask writers to analyse the information presented in another source, usually a graph, and to debate a theoretical question in two separate tasks (IELTS), or to analyse the information presented in another source to identify how the source argues its case (SATs). An implicit understanding of a general academic context is implied in the instructions that ask for a formal style, remind test takers to focus 'on the most relevant features of the passage,' and ask them to analyse how the author of the given text uses language to persuade an audience (College Board n.d. c). In other words, test takers are expected to understand what is required from an academic piece of writing that is situated in a communicative context where the only readers will be the markers and the only purpose for writing is to pass the specific test.

The two examples of large scale writing tests discussed above could thus be seen to assess how likely it is that potential students will be able to pass written assessments in an English speaking HE system. For a test that is used to screen potential students, such a 'measurement [can be said to] to have currency' (Kirk and Miller 1986: 21), as the situation in the text resembles that of many assessments at university. Nonetheless, it is questionable whether this specific definition of 'writing' is the one underlying our teaching of writing, and this is an important question when searching for valid tools that can be used to assess the impact of our practice. Not considering the implications of theoretical assumptions can often be a blind spot in research that uses quantitative tools, as Kirk and Miller identify in their classic volume on qualitative research: 'the survey researcher who discusses attitudes is not wrong to do so [but it is wrong if] he or she fails to acknowledge the theoretical basis on which it is meaningful to make measurements of such entities and to do so with survey questions addressed to a probability sample of voters' (1986: 15). In analogy, it can be argued that it is not necessarily wrong to measure the ability to write a specific form of academic text in test situations with a large scale test that mirrors similar test situations. It should be acknowledged that although even those involved in developing large scale writing tests admit that authentic meaning making is not produced under 'the best social conditions under which to administer standardised tests' (Deane et al. 2008: 65), but this debate is beyond the focus of this paper. It can firmly be asserted, however, that it would be ill-advised to use these tests to assess whether students' ability 'to write' has improved after teaching that defines writing in a very different way.

Clarifying the Notion of 'Writing' in our Teaching

This then means that the search for appropriate tools to assess the impact of our teaching needs to begin with a definition of the construct we are teaching. As indicated above, an academic literacies approach considers writing a 'cultural and social practice' (Lea and Street 1998: 158). In other words, it emphasises that writing is situated and the production of texts, in analogy to Fairclough's definition of critical discourse analysis, 'should not be artificially isolated from [the] analysis of institutional and discursive practices within which texts are embedded' (1995: 9). This means that successful writing depends on knowledge about text types (Bieber 1989), as much as it depends on knowledge of the uses of these text types in discourse communities. To participate successfully in relevant discourse communities, our students need to learn how to 'act, value, interact, and use language *in synch with or in coordination with* other people' (Gee 1999: 14) in different contexts. Following academic literacies again (Lillis 2001), it is also important to foster students' critical engagement with these discourse communities by 'explor[ing] whose interests these rules [of written communication] serve' (Norton 2000: 15).

This construct of writing thus acknowledges its situated nature. Teaching within an academic literacies framework not only means teaching students about the writing process and the linguistic means at their disposal. It also includes teaching students the tools to analyse 'writing as a socially and culturally embedded practice' (Jones 2004: 255), or, in other words, teaching that increases their meta-cognitive awareness with respect to writing in the hope that this increases their ability to 'ask [...] good questions about writing situations and developing heuristics for analysing unfamiliar writing situations' (Elon University 2015).

By unlocking the tacit knowledge (Jacobs 2005: 484) that shapes written interactions in different discourse communities, an academic literacies approach acknowledges the fact that writing, seen as social interaction through written language, is not a generic skill, but one that is developed in specific contexts. It accepts that writing in a different social context represents a consequential transition, i.e. one which, 'when it is consciously reflected on, [and] struggled with [...] shifts the individual's sense of self or social position' (Moore and Anson 2017: 5). Yet, an academic literacies approach assumes that students' ability to participate successfully in written communication benefits from participating in such critical enquiry into the role writing plays in different discourse communities (Coffin and Donohue 2011: 65), or in other terms treating writing not only as an activity, but as a subject of inquiry (Adler-Kassner and Wardle. 2015) . Such critical enquiry is, therefore, as important as writing practice. It helps students 'draw on existing knowledge, skills, processes and attitudes around writing' (Farrell and Tighe-Mooney 2015: 34) and allows them to master academic writing, as well as the challenges of

workplace writing they will face as graduates (Canton, Govan, and Zahn 2017). As the literature on transfer reminds us, participating in new discourses can still present a challenge (Anson 2016), but in an academic literacies framework, addressing this challenge is an intrinsic aspect of teaching writing in all its complexity.

Translating this pedagogical framework into teaching practice is subject to pragmatic limitations in most institutions, but within these restrictions, an academic literacies-inspired approach to teaching writing means teaching students how they can successfully manipulate language to participate in a complex social interaction. Such an understanding of writing offers a major challenge in terms of assessment.

Alternative Methods of Assessing Improvements in Writing as Social Interaction

Most forms of writing assessment do attempt to test writing ability as a complex construct, or situated practice, but there are some notable exceptions. Wardle and Roozen (2012) make an ambitious suggestion for an alternative approach to assessing the impact of the teaching of writing; their ecology of assessment follows the idea of portfolio assessment, but expands this to 'students' writing development across an expansive ecology of literate activities rather than within any single setting' (2012: 107). While Wardle and Roozen's idea that 'rhetorical dexterity [...] across communities and situations [as] perhaps the most common sort of literate practice' (2012: 107) clearly echoes the construct of writing set out above, their plans require an at least institution-wide approach, as well as significant amounts of funding, which renders the approach little useful in a context where small groups of lecturers or small institutions, such as writing centres, want to evidence the impact of their teaching.

The tradition of portfolio assessment could be seen to provide a smaller and more manageable version of this approach, as the comparison of different texts produced under different circumstances is more conducive to considering writers' ability to respond appropriately to a specific communicative situation. Holistic assessment tends to be classified as 'pass' or 'fail' (Yancey 1999), however, which does not allow a nuanced evaluation of writers' progress. Therefore, this tradition does not offer a suitable, tested measurement tool suitable to evaluate our teaching either.

A more gradual form of assessment of written portfolios, however, could capture progress better, and a search beyond the field of writing suggests a novel approach: adapting a tool from a different field to writing. Amabile's Consensual Assessment Technique (CAT) for creativity (Amabile 1982, Amabile 1996, Hennessey, Amabile, and Mueller 2011) offers a measurement tool that uses expert judgement and consensus, originally to determine the creativity of a product. The following sections of this paper will explore the parallels between creativity, in Amabile's contexts, and writing, as defined in our teaching. The paper then examines the similarities between these two concepts, whose situated nature makes both complex and ill-defined problems (Carey et al. 1989). On this basis, I argue that the principles of Amabile's work could be transferred to the evaluation of written texts, which in turn would allow conclusions about writers' abilities.

Adapting Amabile's Work to Writing

Functional Parallels between Amabile's Concept of Creativity and Writing

Research on creativity might not be the most obvious discipline to consult in search of tools to assess successful writing, but writing and creativity share an important characteristic: the lack of a definitive list of 'specific traits' or 'objectively identifiable features' (Amabile 1996: 19) that remain relatively stable across different contexts. As discussed above, an academic literacies approach considers writing to be situated, which means that it resembles creativity in lacking a 'precise, *universally applicable* definition' (Amabile 1996, 19; emphasis added). This lack makes both concepts difficult to assess, and Amabile's discontent with previous methods of assessing creativity, based on specific characteristics, led her to develop the CAT as an alternative measurement tool. Her point of departure is not dissimilar to the rationale for this research project, outlined in the first part of this paper.

The fact that a concept is not easily defined or stable across contexts might be an incentive to search for alternative forms of measurement, but it does not guarantee that this is possible. The feasibility of scientific scrutiny of such apparently fuzzy and elusive concepts (Amabile 1996:19) rests on two main criteria, which Amabile identifies as prerequisites for the use of CAT. As the following paragraphs demonstrate, both prerequisites are fulfilled by the concept of successful writing that appropriately responds to specific communicative situations and communicates its purpose well to its readership.

The first criterion Amabile establishes is the fact that creativity is something many people feel they 'can recognise and agree upon when they see it' (Hennessey, Amabile, and Mueller 2011: 253), even if they cannot name its specific characteristics. This does not mean that creativity cannot be controversial, or that there can be no disagreement over the creativity of products, but it recognises that in many contexts people are happy to act based on the assumption that there can be general agreement on creativity.

The same could be argued for successful writing: it is by no means an unproblematic concept, and disagreements about what constitutes the details of successful writing are common, but the notion that 'good' writing can be distinguished from 'bad' writing underlies many everyday practices. It is indirectly evident, for example, in numerous complaints about the lack of writing skills in different groups. Natalie Wexler, for example, suggests, 'any college professor or employer' could confirm that 'many Americans are lousy writers' (2015). Her complaint faithfully reports those that have been made elsewhere over a long period of time. In 1964 Fielden expressed similar concerns about managers in *The Harvard Business Review*; in 2006 the Royal Literary Fund 'communicate[d] grave concerns about shortcomings in student writing skills nationally' (2006: vii) in the UK; and in Germany 'university lecturers are raising the alarm [because] many prospective teachers cannot put two sentences on paper without making mistakes' (Hock and Onkelbach 2015; translation mine). All these sources suggest the common perception that good writing can be recognised, at least by experts who carry authority in their fields, such as the higher-ranking managers cited by Fielden who lose a significant amount of time 'having to do reports and letters [written by their staff] over and over before they go out' (1964). Other groups whose knowledge of the standards for good writing is often cited, or expressed in the writing advice they share with the world, are university lecturers, teachers, journalists or professional writers, in other words, those whose authority rests on greater practice and experience in writing, or on social recognition for their own texts. The tacit assumption that these groups are competent judges of the quality of writing and, more importantly, that they more often than not agree in their judgements, suggests that the success of writing is less problematic in everyday situations than in scientific enquiry. A tool like the CAT, which could systematically explore this consensus, would allow us to explore the extent to which these tacit assumptions withstand more rigorous testing.

In addition to implicit references to the intuitive expertise of specialists, writing is also explicitly regarded as a means to an end, which is successful if it fulfils its function. This function can be relatively imprecise, as in a writer's advice that 'analysing the stuff I read once in a while to figure out why they work (or don't) and how they made something work (or didn't) is extremely useful' (Wrede 2010), or very precise, as in the Preface to the US Army Writing Manual where the Chief of Staff stipulates that good writing needs to ensure that 'soldiers and civilians two echelons removed will know the end we seek' (Department of the Army 1986) and carry out the intended order as a result. In both instances, successful writing is not characterised by specific features, but by its function. There might be some disagreement over these functions and the level to which specific texts fulfil them, but in everyday situations evaluations of writing based on tacit assumptions is common, and more often than not, little problematic. It can, therefore, be expected that some consensus over this concept exists and can be measured.

The second criterion required to make something assessable through consensus is that it can be considered to come in degrees. Most people agree that some products are more or less creative than others (Amabile 1996: 34), and the same can be argued in terms of writing. The notion that some texts are more or less appropriate in a specific situation is rather intuitive. In the context of this second criterion, Amabile further specifies the perspective from which the judgements on the degree of creativity are made: although evaluations of a product's creativity

are invariably based on the observers' previous experience and intuitive standards, they will consider whether the task of creating it was 'heuristic for the individual in question' (1996: 36). The person assessing creativity thus takes into account the perspective of the person who created a product.

This is similar again to the way readers draw conclusions about texts: while they compare them to their understanding of the communicative context and their previous experiences with written texts, their evaluation will also be based on their interpretation of the writer's perspective. Amabile exemplifies this point with the fact that raters will rate children's products differently from the way they will rate adults', and this will apply in the same way to text evaluation (Amabile 1996: 36). For writing, readers also seem to take consideration the writer's communicative purpose and experience when evaluating a text. Anecdotal evidence from education suggests that teachers' and lecturers' expectations of students' writing seem to be directly influenced by students' age and experience, whether these expectations are confirmed by official curricula or not. The persuasive arguments pupils are asked to produce at Scottish primary level 4, for example, would be considered to be very weak writing for a level one undergraduate student. Similarly, many lecturers affirm that their expectations for writing at Master's level is different from those they have for undergraduate students, a distinction that is, ironically, taken up prominently by companies that help students cheat with custom-written essays (Papermasters n.d.). The prevalence of these considerations is further confirmed by structures outside the education system that differentiate between writers of different ages, such as writing competitions for children that specify eligible age groups (Tolkien 2013), training opportunities for younger adults (e.g. under 35 Second Opportunity Africa 2017), or opportunities to publish that are aimed at young writers (Aerogramme Writers' Studio 2015).

As the previous paragraphs have demonstrated, the notion of successful writing fulfils similar criteria to creativity, in terms of how writing is operationalised and evaluated in practice beyond scientific enquiry. Consequently, Amabile's operational definition for the CAT – that 'a product or response is creative to the extent that appropriate observers independently agree it is creative' (1982: 1001) – can be applied to writing. An operational definition of successful writing, within the context of an academic literacies approach, could therefore be as follows:

A written text can be considered successful to the extent that appropriate observers independently agree it is successful within a specific communicative situation.

It could further be expected that Amabile's approach, i.e. to develop a measurement tool that is based on such an operational, or functional, definition, can be transferred to the concept of successful writing, and that such a tool would produce a similar level of consensus among raters as the studies presented in Amabile (1996).

Further Arguments for the Appropriateness of CAT for Writing

While Amabile declares the existence of an operational definition based on consensus sufficient for the scientific study of creativity, she also stresses the importance of using a working conceptual definition. For her, it is important to 'build [...] a theoretical formulation of the creative process' (1996: 33) in order to advance the study of creativity. Contributing further to the study of writing is not the primary aim of this research project, partially because numerous disciplines, ranging from cognitive psychology to applied linguistics, discourse analysis and other areas of linguistics, already offer a wealth of research into the concept, process and products of writing. The focus here is on finding a solution for the practical, *pedagogical problem* of how we can evaluate the impact of our teaching. Searching the existing body of literature on writing for a conceptual definition that supports the operational definition above can still be very useful, however: if it is possible to find a conceptual definition that both supports the operational definition *and* is rooted in existing knowledge about writing, the validity of this assessment tool increases further. This would show the tool is not just valid to the construct it claims to measure (i.e. the operational definition), but also valid in terms of the previously developed conceptual understandings of this construct.

Such a conceptual definition can be found in Fairclough's (1995) conception of the centrifugal and centripetal pressures that shape textual production. For Fairclough, centrifugal pressures

arise from the need to use language to generate new meanings, and modify text types and discourses. These are countered by centripetal pressures borne from the need to draw upon conventional use of linguistic means to ensure mutual understanding (1995: 7-8). Successful writing achieves a good balance between these forces by keeping within the limits of comprehensibility while using language in new ways to allow for nuanced, individual and innovative ideas to be expressed. Fairclough's notion echoes Amabile's conceptual definition of creative products as those that constitute 'both a novel and appropriate, useful, correct or valuable response to an open-ended task' (Hennessey, Amabile, and Mueller 2011: 253), where novelty (centrifugal) needs to be in balance with the restraints of usefulness or appropriateness (centripetal).

Fairclough also stresses that the balance between centrifugal and centripetal forces in (written) language depends on the historical, social and cultural contexts. These contexts determine first, what expected conventional uses of linguistic means are; and second, how far individual meaning making can depart from these conventions to express original ideas, while still remaining comprehensible. Here again the concept of writing as an ongoing process of negotiating the ideal balance between originality and comprehension reflects Amabile's assertion that 'criteria for creativity require an historically bound social context' (1996: 34). These parallels between Fairclough's and Amabile's conceptual definitions offer further support to the argument that the CAT can be transferred to the assessment of good writing as successful communication.

Applying the CAT to Writing

Together these conceptual parallels provide significant support to the idea that the concept of assessing creativity as a recognisable, gradual quality can be successfully applied to the field of writing. In an operational definition, successful writing can be defined through readers' consensus on the degree to which a text fulfils its communicative purpose, a definition that includes both reader and writer. Assessing the impact of our teaching based on this definition would also address the problem of conceptual validity, which arises for most large scale writing tests whose definition of writing does not reflect the one we use in our teaching. Amabile explains that in the CAT construct validity can be established through high inter-rater reliability (Hennessey, Amabile, and Mueller 2011: 256). The operational definition of creativity claims that the concept of creativity exists, because it is possible to agree on the level of creativity of products, not necessarily because these products share specific characteristics. If this agreement is empirically supported through high inter-rater reliability, their consensus then supports the operational definition as well, suggesting creativity is a valid construct (Amabile 1982: 1002). If a similar level of inter-rater reliability could be achieved on the communicative appropriateness of texts, it could be demonstrated that successful writing is another concept well-defined by function, even if the characteristics are hard to isolate. This means that, while the conceptual parallels may be convincing, the feasibility of applying the notion of consensual assessment to writing can only be confirmed through empirical work which produces high inter-rater reliability.

Grounding the potential for this new measurement instrument requires further empirical testing. An initial pilot study has been conducted, with limited success, and a second study addressing the pilot's shortcomings is currently under way. Although this paper focuses on laying the theoretical foundations of such empirical studies, I will finally sketch out the different steps and considerations underpinning the empirical research.

First of all it is necessary to design rating scales based on the operational definition that 'a written text can be considered successful to the extent that appropriate observers independently agree it is successful within a specific communicative situation. This raises further questions about the type and scope of the scale used. Rating scales used in Amabile's CAT studies contained the question of how creative a product is, as well as additional items. These were included to test whether creativity was evaluated as an independent construct, since high correlations between the item asking for creativity and items focusing on other factors could indicate unwanted confusion between the different constructs measured.

A CAT-based writing measurement tool would need to determine whether the success or appropriateness of a text can be evaluated with a single question, and whether to include further items to test the independence of this construct. In addition the type of response needs to be chosen. Amabile (1996) uses a scale from 1-10, but research in other areas that measure subjective impressions, such as pain research (Couper et al. 2006), or food perception research (Lawless and Heymann 2010: 155), suggests visual analogue scales or line scales can provide a potential alternative. Once the rating scale has been designed, it is necessary to identify texts produced for the same, or at least a highly similar, communicative situation, which can be evaluated by 'appropriate observers' (Amabile 1982: 1001). Neither Amabile nor her collaborators on CAT stipulate a minimum number of products, but for statistical testing the number of texts to be evaluated should not be too small. After using 7 texts for the pilot study, the current study has expanded to 29 texts to examine whether some of the problems resulted from a smaller number of texts, but further research is needed to determine a benchmark.

Defining the number and criteria for these suitable raters is a significant challenge. Hennessey, Amabile, and Mueller (2011: 265) report that many of their studies used approximately ten raters, which exceeds the number of colleagues available for such a task in many writing and learning support centres, for example. Initially, Amabile restricted the group of appropriate raters to 'those familiar with the domain in which the product was created or the response articulated' (1996: 33). Later studies in which the CAT was used for creative products, however, achieved high inter-rater reliability for non-expert groups as well (1996: 72). Arguably, this was possible because most CAT studies were deliberately chosen not to require any specific skills, ensuring the measurement of creativity was not confounded by technical ability.

Writing, on the other hand, is strongly based on specialised skills, and in the context of evaluating the teaching of writing, it can even be argued that the main objective is to measure any changes in these skills. This difference between the two constructs needs to be carefully considered, as it can have important implications for the choice of appropriate raters. In addition to thinking about how much 'intuitive expertise' (Amabile 1996: 42) raters need to have developed in terms of successful communication, it is also necessary to take into account the framing of the rating task, since the way in which the communicative situation is presented is likely to determine the implicit standards raters will use in their evaluations. Although the interaction between these implicit standards and the process of comparing the different products to be rated is not entirely clarified in Amabile's work, it is certain that both of these points of comparison influence raters' decisions. As a result, it is important to consider how the presentation of the communicative context can influence the degree to which raters rely on each of these processes, and which aspects of their intuitive expertise they draw upon.

Conclusion

If empirical studies confirm the CAT can provide reliable judgements on the success of writing, it would become possible to use this form of evaluation across different contexts without defining the specific characteristics of good texts for each specific situation. This would potentially make it possible to compare different students or texts produced by the same student, for example, before and after teaching interventions that aim to increase students' ability to manipulate written language to achieve their communicative purposes. An increase in the level to which their texts achieve their communicative purpose, or in their communicative performance (Weigle 2002: 46), could be seen as a reflection of increased writing ability. Such an assessment might still not include the critical engagement with language we want to foster in an academic literacies approach, but it could provide evidence that our teaching has an impact on students' ability to adapt their writing appropriately to specific communicative situations, assuming this is partly based on their ability to analyse these situations. .

This paper has demonstrated the need for a significant amount of further research, to empirically test whether applying consensual assessment to the construct of successful writing is possible. The list of factors presented above that need to be considered for developing such empirical research indicates this is a time consuming undertaking. Even if studies could achieve high inter-rater reliability, and thus prove consensus on what makes written communication

successful exists, further research would still be needed to examine whether measurement instruments based on consensus are sufficiently sensitive to identify changes in writers' performance or not. In other words, the suggestion that the CAT has the potential to develop into a measurement tool that helps us evaluate the impact of our teaching is still a tentative one, and even if its potential is fulfilled, it could still constitute a time consuming, resources-intensive method.

Nonetheless, it can be argued that it is worth pursuing this line of enquiry. An application of the CAT to writing could address many of the problems associated with current methods of evaluating writing support and the teaching of writing outlined in the first part of this paper. Consensual assessment would redirect the focus from the writer to the interaction between reader and writer. Defining the success of writing as successful interaction would also avoid the problems with construct validity that arise for most large scale tests, as it is more compatible with our conceptual understanding of writing as social interaction, which is, by definition, subjective and evaluated by recipients. It would do justice to the construct we are teaching, and assess it in a range of contexts that can reflect those in which our students are required to write, rather than in artificially created test situations. Through this, consensual assessment of writing could help us counter calls to resort to overly simple forms of measuring our impact. At the same time, it would be able to deliver results based on a formal method of assessment that goes beyond anecdotal evidence. In addition to the benefits this would offer us as a profession in terms of accountability, it would open up new possibilities for evidence-based practice, since it would create an additional, highly valuable form of feedback on our teaching. These potential benefits seem to warrant further research into consensual assessment as a novel approach to the evaluation of writing.

Acknowledgements

The author would like to thank Morag Ferguson who recommended Amabile's work to her and Daniela Zahn whose constructive criticism was invaluable throughout the many revisions this paper has undergone.

References

- ACT (n.d.) *Writing Test Description*. [online] available from
<<http://www.actstudent.org/testprep/descriptions/writingdescript.html>> [03 August 2018]
- Adler-Kassner, L. and Wardle, E. (2015) *Naming what we Know*. Logan: Utah State University Press
- Aerogramme Writer's Studio (2015) 'Ten Publication Opportunities for Young Writers.' *News and Resources*. [online blog entry] 26 March. available from
<<https://www.aerogrammestudio.com/2015/03/26/publication-opportunities-for-young-writers/>> [8 August 2018]
- Ahmad, R. and McMahon, S. (2006) 'The Benefits of Good writing.' in *Writing Matters*. ed. by Davies, S., Swinburne, D. and Williams, G. [online] available from
<<http://www.rlf.org.uk/fellowshipscheme/research.cfm>> [03 August 2018]
- Amabile, T. M. (1982) 'Social Psychology of Creativity: A Consensual Assessment Technique.' *Journal of Personality and Social Psychology* 43 (5), 997–1013
- Amabile, T. M. (1996) *Creativity in Context*. Boulder, CL: Westview Press
- Anson, C. M. (2016) 'The Pop Warner Chronicles: A Case Study in Contextual Adaptation and the Transfer of Writing Ability.' *CCC* 67 (4), 518-549
- Banerjee, J. and Wall, D. (2006) 'Assessing and Reporting Performances on Pre-Sessional EAP Courses: Developing a Final Assessment Checklist and Investigating its Validity.' *Journal of EAP* 5, 50-69
- Bharutram, S. and McKenna, S. (2006) 'A Writer–Respondent Intervention as a Means of Developing Academic Literacy.' *Teaching in Higher Education* 11 (4), 495-507
- Bieber, D. (1989) 'A Typology of English Texts.' *Linguistics* 27, 3-43
- Canton, U., Govan, M. and Zahn, D. (2017) 'Rethinking Academic Literacies. A Conceptual Development based on teaching practice.' *Teaching in Higher Education* [online] available from
<<http://www.tandfonline.com/doi/full/10.1080/13562517.2017.1414783>> [03 August 2018]
- Carey, L., Flower, L., Hayes, J. R., Schriver, K. A., Haas, C. (1989) 'Differences in Writers' Initial Task Representations.' *Technical Report*. Centre for the Study of Writing and Carnegie Mellon Planning Project [online] available from
<https://www.nwp.org/cs/public/download/nwp_file/78/TR35.pdf?x-r=pcfile_d> [03 August 2018]
- Coffin, C. and Donohue, J. P. (2011) 'Academic Literacies and systemic functional linguistics: How do they relate?' *Journal of English for Academic Purposes* 11, 64-75
- College Board (n.d.a) *Frequent Questions about SAT*. [online] available from
<<http://sat.collegeboard.org/about-tests/sat/faq>> [03 August 2018]

- College Board (n.d.b) *SAT Essay Scoring*. [online] available from
<<https://collegereadiness.collegeboard.org/sat/scores/understanding-scores/essay>>
[03 August 2018]
- College Board (n.d. c). *SAT Practice Essay # 2*. [online] available from
<<https://collegereadiness.collegeboard.org/pdf/sat-practice-test-2-essay.pdf>> [03
August 2018]
- Couper, M. P., Tourangeau, R., Conrad, F. G., and Singer, E. (2006). 'Evaluating the
Effectiveness of Visual Analogue Scales. A Web Experiment.' *Social Science
Computer Review* 24 (2), 227-245
- Deane, P., Odendahl, N., Quinlan, T., Folwes, M., Welsh, C., and Bivens-Tantum, J. (2008)
Cognitive Models of Writing: Writing Proficiency as Complex Integrated Skills. [online]
available from
<http://www.ets.org/research/policy_research_reports/publications/report/2008/htxo>
[03 August 2018]
- Department of the Army (1986) *Effective Writing for Army Personnel*. Pamphlet 600-67.
[online] available from <http://www.au.af.mil/au/awc/awcgate/army/p600_67.pdf> [03
August 2018]
- Diamond, A., Walkely, L., Forbes, P., Hughes, T. and Sheen, J. (2011) *Global Graduates into
Global Leaders*. [online] available from
<[http://www.ncub.co.uk/index.php?option=com_docman&view=download&category_s
lug=publications&alias=42-global-graduates-into-global-leaders&Itemid=2728](http://www.ncub.co.uk/index.php?option=com_docman&view=download&category_slug=publications&alias=42-global-graduates-into-global-leaders&Itemid=2728)> [03
August 2018]
- EATAW Members (2017) 'Material from Workshop "Evaluating Writing and the Teaching of
Writing". Session 14 conducted by Canton, U.' *EATAW Conference*. Royal Holloway,
London. [19 – 21 June, 2017]
- Elon University (2015) *Elon Statement for Writing Transfer*. [online] available from
<[http://www.elon.edu/e-
web/academics/teaching/ers/writing_transfer/statement.xhtml](http://www.elon.edu/e-web/academics/teaching/ers/writing_transfer/statement.xhtml)> [03 August 2018]
- Fairclough, N. (1995) *Critical Discourse Analysis: The Critical Study of Language*. Harlow:
Longman
- Farrell, A., and Tighe-Mooney, S. (2015) 'Recall, Recognise, Re-Invent: The Value of
Facilitating Writing Transfer in the Writing Centre Setting.' *Journal of Academic
Writing* 5 (2): 29-42
- Fielden, J. S. (1964) 'What do you Mean I can't Write?' *Harvard Business Review*. [online]
available from <<https://hbr.org/1964/05/what-do-you-mean-i-cant-write>> [03 August
2018]
- Gee, P. (1999) *An Introduction to Discourse Analysis: Theory and Method*. London:
Routledge
- Hendry, G. D., and Dean, S. J. (2002) 'Accountability, evaluation of teaching and expertise in
higher education.' *International Journal for Academic Development* 7 (1), 75-82.
[online] available from
<<https://www.tandfonline.com/doi/abs/10.1080/13601440210156493>> [03 August
2018]

- Hennessey, B. A., Amabile, T.M., and Mueller, J. S. (2011) 'Consensual Assessment.' in *Encyclopaedia of Creativity*. ed. by Runco, M. A. and Pritzker, S. R. Amsterdam: Academic Press / Elsevier, 253-60
- Hook, S., Onkelbach, C. (2015) 'Dozenten klagen: Lehramtsstudenten können nicht schreiben'(Student Teachers cannot Write) *Der Westen*. [online] available from <<https://www.derwesten.de/politik/dozenten-klagen-lehramtsstudenten-koennen-nicht-schreiben-id10441860.html>> [03 August 2018]
- IELTS (n.d.a). *IELTS Assessment Criteria*. [online] available from <<http://takeielts.britishcouncil.org/find-out-about-results/ielts-assessment-criteria>> [03 August 2018]
- IELTS (n.d.b). *IELTS Sample Test Questions*. [online] available from <<https://www.ielts.org/about-the-test/sample-test-questions>> [03 August 2018]
- Jacobs, C. (2005) 'On being an insider on the outside: new spaces for integrating academic literacies.' *Teaching in Higher Education* 10 (4), 475 – 87
- James, D. (2005) 'Importance and impotence? Learning, outcomes and research in further education.' *The Curriculum Journal* 16 (1), 83-96 [online] available from <<https://www.tandfonline.com/doi/abs/10.1080/0958517042000336827>> [03 August 2018]
- Jones, J. (2004) 'Learning to Write in the Disciplines.' in *Analysing Academic Writing: Contextualised Frameworks*. ed. by Ravelli, L., Jones, J., and Ellis, R. A. London: Continuum
- Kirk, J., and Miller M. L. (1986) *Reliability and Validity in Qualitative Research*. Beverly Hills: Sage
- Lawless, H. T., and Heymann, H. (2010) *Sensory Evaluation of Food: Principles and Practices*. 2nd edn. New York: Springer
- Lea, M. R., and Street, B.V. (1998) 'Student Writing in Higher Education: An Academic Literacies Approach.' *Studies in Higher Education* 23(2), 157-172. [online] available from <<http://dx.doi.org/10.1080/03075079812331380364>> [03 August 2018]
- Leibowitz, B., Goodmann, K., Hannon, P., and Parkerson, A. (1997) 'The Role of a Writing Centre in Increasing Access to Academic Discourse in a Multilingual University.' *Teaching in Higher Education* 2 (1), 5-19
- Li, L. Y. and Vandermensbrugge, J. (2011) 'Supporting the Thesis Writing Process of International Research Students through an Ongoing Writing Group.' *Innovations in Education and Teaching International* 48 (2), 195-205
- Lillis, T. M. (2001) *Student Writing: Access, Regulation, Desires*. London: Routledge
- Moore, J. L., and Anson, C. M. (2017) 'Introduction.' in *Critical Transitions: Writing and the Question of Transfer*. ed. by Anson, C. M., and Moore, J. L. Fort Collins: WAC Clearinghouse and University Press of Colorado, 3-13
- Norton, B. (2000) *Identity and Language Learning: Gender, Ethnicity and Educational Change*. Harlow: Pearson

- Nzekwe-Excel, C. (2014) 'Academic writing workshops: Impact of attendance on performance.' *Journal of Academic Writing* 4 (1), 12– 25
- Papermasters.com (n.d.) 'Why Graduate Level Research Projects Need Custom Writing.' *Papermasters.com* [online] available from <<https://www.papermasters.com/graduate-research.html>> [8 August 2018]
- Regan, J. A. (2012) 'The Role Obligations of Students and Lecturers in Higher Education.' *Journal of Philosophy of Education* 46 (1), 14-24
- The Royal Literary Fund (2006) *Writing Matters*. ed. by Davies, S., Swinburne, D., and Williams, G. [online] available from <<https://www.rlf.org.uk/wp-content/uploads/2014/05/RLFwritingmatters.pdf>> [03 August 2018]
- Second Opportunity Africa Foundation (2017) 'British Council Young Critics Programme 2017'. [online] available from <<https://secondopportunityafrica.org/british-council-nigeria-young-critics-programme-2017-for-young-writers-journalists/>> [8 August 2018]
- Sengupta, S., and Leung, K. (2002) 'Providing English Language Support through Collegial Mentoring: How do we Measure its Impact?' *Assessment and Evaluation in Higher Education* 27 (4), 365-81
- Thonus, T. (2002) 'Tutor and Students Assessments of Academic Writing Tutorials: What is "Success"?' *Assessing Writing* 8, 110-34
- Tolkien, T. (2013) 'Children's Writing Competition'. *The School Reading List*. [online blog entry] 19 May. available from <<https://schoolreadinglist.co.uk/competitions-for-children/childrens-writing-competitions/>> [8 August 2018]
- Wardle, E., and Roozen, K. (2012) 'Addressing the Complexity of Writing Development: toward an Ecological Model of Assessment.' *Assessing Writing* 17, 105–119
- Weigle, S. C. (2002) *Assessing Writing*. Cambridge: Cambridge University Press
- Wexler, N. (2015) 'Why Americans Can't Write.' *The Washington Post* [online] 24 September. available from <https://www.washingtonpost.com/opinions/why-americans-cant-write/2015/09/24/6e7f420a-6088-11e5-9757-e49273f05f65_story.html> [03 August 2018]
- Wind, S. A., and Engelhard Jr, G. (2013) 'How Invariant and Accurate are Domain Ratings in Writing Assessment?' *Assessing Writing* 18, 278-99
- Wrede, P. C. (2010) 'Recognising 'Good Writing.' *PC Wrede: 6 Impossible Things*. [online blog entry] 6 June. available from <<http://www.pcwrede.com/recognizing-good-writing>> [03 August 2018]
- Yancey, K. B. (1999) 'Looking Back as we Look Forward: Historicising Writing Assessment.' *College Composition and Communication*. 50 (3), 483–503