A dual purpose data base for research and diagnostic assessment of student writing

Judy M. Parr

University of Auckland | New Zealand

Abstract: The data base of writing examined serves a dual purpose. Here it is used as a research tool and the writing performance from the large, nationally representative sample (N = 20,947) of students (years 4 to 12) interrogated to examine patterns of performance in writing. However, the data base was designed to underpin a software tool for diagnostic assessment of writing. Viewing writing as accomplishing social communicative goals, performance was considered in terms of seven main purposes the writer may seek to achieve. Tasks related to each purpose were encapsulated in 60 writing prompts that included stimulus material. Participants produced one writing sample; the design ensured appropriate representation across writing purposes. Samples were scored using criteria differentiated according to purpose and curriculum level of schooling and acceptable reliability obtained. Analyses indicate that growth was most marked between years 8 and 10, arguably, as opportunity to write increases and writing is linked to learning in content areas. Variability in performance is relatively low at primary school and high at secondary school. Students at any level did not write equally well for different purposes. Mean scores across purposes at primary school were relatively similar with to instruct and to explain highest. By years 11-12 there is a considerable gap between the highest scores (for narrate and report) and the lowest, recount, reflecting likely opportunities to practice writing for different purposes. Although girls performed better than boys, the difference in mean scores narrows by years 11-12.

Keywords: composition, writing development, genre

journal of WRITING RESEARCH

Parr, J. (2010). Exploring the Potential of Interrogating a Mediated Corpus of Data From a National Representative Sampling of Student Writing. *Journal of Writing Research, 2* (2), 129-150. http://dx.doi.org/10.17239/jowr-2010.02.02.3

1. Introduction

Corpora are valuable tools in understanding language use and patterns in the development and acquisition of language. A corpus is generally seen to be a collection of oral or written texts input into electronic form so that they are accessible for analysis. Traditionally oral language was captured and analysed in this fashion but the electronic age has made it easier (and cheaper) to compile certain types of written rather than spoken corpora. Written texts appear in general corpora. The Brown corpus (Kucera & Francis, 1967) has a category "Learned" consisting of 80 written texts, while the British National Corpus (1994) contains academic texts and the Bank of English corpus has academic text books (Krishnamurty & Kosem, 2007). Corpora have been devised for research purposes and for pedagogical purposes; technological progress has meant that corpora may be collected for personal research or teaching. Some are large data bases, others small, in-house collections. Some designed for one purpose, are utilised for another and, as this article illustrates, a corpora designed primarily to support one purpose, can also provide a rich source of data for the other. The data base of writing that is the subject of this article serves a dual purpose. As a research tool, it can be interrogated regarding patterns of written language but it was primarily designed as the underpinning data base for a software tool for diagnostic assessment of writing.

The availability of corpora of texts has fostered a quantitative view of the patterns of language usage, often focusing on frequency based and/or distributional analyses of psycholinguistic phenomenon. For research, corpora facilitate the testing of theories concerning language like, for example, about the extent to which structural frequency plays an explanatory role in various psycholinguistic phenomena. Recent models of language comprehension emphasise the role of distributional frequency of words and structures, along with a user's experience, in relation to the accessibility or ease of processing of a particular lexical item or sentence (Roland, Dick & Elman, 2007) and corpora allow these models to be both built and tested. A recent example involves an analysis of an individual performing the complex task of talking in a foreign language. Here the link between lexical competence and oral fluency has been investigated using a corpus of oral productions in three different L2s (Hilton, 2008). Written corpora have seen the reappearance of vocabulary frequency profiling and word lists. This has allowed an examination of breadth and depth of vocabulary in relation to writing (Laufer & Nation, 1995). Written corpora have been used to determine the features of an academic register, largely in terms of word frequency and specific vocabulary use (Coxhead, 2000). Notably, these analyses have involved a narrow set of structures; corpora have had only a small role in research on higher level language processing (Roland, Dick & Elman, 2007).

Regarding pedagogical purposes, there is a large amount of research into using corpora for language learning in general (Braun, Kohn & Mukherjee, 2006) and particularly in relation to the use of language for academic purposes (e.g. Krishnamurty & Kosem, 2007). The argument is that corpora can provide an open-ended source of

language data, tailored to a learner's needs and promote a learner-centred, discovery approach to learning (Hyland, 2006). With respect specifically to their use in writing, one use seems to be as models, particularly for foreign language learners. Rohrbach (2003, cited in Braun, 2007), for example, used a small corpus of tourist brochures written in English to sensitise German speakers to the discourse moves and the means of expression associated with them in such material. Corpora can be used by students to check English usage, for example, when writing a text and for learning grammatical rules (Vannestal & Lindquist, 2007). A comparison of learner corpora and native speaker corpora point to common errors in writing made by non native speakers and could clearly be used to inform teaching practice. However, a discussion of a different type of pedagogical use of corpora, namely, in terms of teachers learning about their students or about likely patterns of learning in their students, is largely absent from the literature.

There are two significant, inter-related issues, one theoretical and one methodological, that arise in the use of corpora of written texts whether for research or pedagogical purposes. Then, there is a third, pragmatic issue in terms of obtaining the writing of school students. With respect to the theoretical considerations, there is a fundamental disjuncture between the use of corpora, however rich, as exemplifying texts and the social communicative nature of language and of language learning (Braun, 2007). With respect to psycholinguistic features, arguably, different genres of language and different contexts of use give rise to different probabilities for the same structures (Biber, 1993; Merlo, 1994). In written text, the production of different texts by the same writer may vary in features and quality (Hayes, Hatch & Silk, 2000) and the features and quality of an individual's production of the same genre may vary across contexts. This also has methodological implications in that selections have to be made both in obtaining the data for the corpora and, in turn, from the corpora for processing in order to analyse certain linguistic features. The second issue is that to obtain, for example, the relative frequency of sentence structures, the data have to be processed, generally parsed, in order to identify syntactic structures. This process is plagued by difficulties in obtaining parser reliability and also transparency of the extraction process (Roland, Dick & Elman, 2007). With respect to the nature and quality of writing there are comparable issues around reliability of judgements in writing assessment (Williamson & Huot, 1993). Finally, with regard to building corpora of the writing of young people, there are logistical issues. Published texts are relatively easy to acquire while acquiring student texts is acknowledged to be more time consuming and potentially frustrating (Nesi et al, 2005). A key issue with the writing of younger students is that it is generally not available electronically; the large part of school student work, particularly in the primary/ elementary school years, is still handwritten.

Each of the above issues was a consideration in this paper. Given the difficulty of obtaining representative samples of the work of school students in electronic form, this paper adopts a more encompassing view of corpora. While the corpus under discussion contains data from the writing samples of a large number of students aged 9 to 17, it is

the results of mediation of the actual handwritten texts (in this case through detailed criterion scoring) that are entered into the data base in order to be analysed. In linguistics there is talk of a corpus-informed approach to answering research and pedagogical questions. This approach is one that allows the linguist to "mediate the corpus, design it from the very outset and build it with applied linguistic questions in mind, ask of it the questions applied linguists want answers to, and filter its output, use it as a guide or tool for what you the teacher want to achieve" (McCarthy, 2001, p.129). The notion of building a corpus to answer specific questions that are of interest to teachers applies to the mediated corpus of writing data discussed in this article; the corpus is the normative data that underpin a sophisticated diagnostic tool for teachers in writing (www.asTTle.org)[1] allowing them to ask questions about their own students' performance. However, although gathered in the service of diagnostic assessment, this data base is large and representative so also allows a number of potentially informative research questions to be asked of it concerning patterns of development. The results of asking research questions of the corpus of writing data form the major part of the present article.

In arriving at the mediated corpus, several theoretical (and methodological) issues had to be considered. The issue that writing serves a social communicative function and that performance may vary across purposes and contexts was, in part, addressed by the theoretical stance underpinning the design of the diagnostic tool and also, in part, addressed by the sampling procedures for the normative sample. While the sampling design was not repeated or longitudinal, the systematic sampling of a large number of students ensured that it was representative of both students nationally and of a range of communicative purposes (genres) for writing. The notion of communicative purposes for writing is the view of writing encapsulated in the design of the procedures used initially to generate the corpus of data about writing performance and, subsequently, in the tool designed for diagnostic assessment that contained as a necessary and integral part of it, the normative performance data. The theoretical stance is that writing is a social and cultural practice, where the term 'genre' refers to the processes involved in 'getting things done' through language (Kress, 1993). A text can be seen from two perspectives: "a thing in itself that can be recorded, analysed and discussed, and also a process that is the outcome of a socially produced occasion" (Knapp &Watkins, 2005, p. 13). Forms of text produced (i.e. genres) in and by specific social institutions (like within schooling) will have some stability, to the extent that there is relative stability of the social structures (Kress, 1993). Given this relative stability, the work of the functional genre theorists who identified common genres or patterned responses (e.g. Martin, Christie & Rothery, 1987) was also relevant.

2. Development and Use of the Diagnostic Tool

In the design of the data base and diagnostic tool, writing was seen as serving seven major purposes, a core set of generic processes that encapsulate what the text is doing

(Knapp & Watkins, 2005). These purposes are: (i) to inform or entertain through narrating or "storying" (imaginative narrative, personal interpretive/ expressive); (ii) to inform or entertain through recount; (iii) to report and describe by classifying and organizing information; (iv) to instruct or lay out a procedure; (v) to argue or persuade; (vi) to explain (after Knapp & Watkins, 1994) and (vii) to analyse.

For each of the major purposes that inform, or processes that form texts, an analytic rubric was developed. Descriptions of features and text structures *commonly* associated with a generic social purpose were utilized to inform the criteria developed in the rubrics. Using the professional expertise of the designers and informed by the work of others (e.g. Derewianka, 1990; Wing Jan, 1991), for each communicative purpose, specific criteria were devised for the dimensions of audience awareness and purpose, content and ideas, structure and language resources, while the criteria for grammar, spelling and punctuation were the same across the seven purposes. The criteria were differentiated by levels of the national curriculum which, in this tool, covers levels 2 to 6 which apply to years 4 to 12 of schooling (ages 9 to 17 approximately). Appendix A contains the analytic rubric for the purpose 'to persuade'. (Note: Only the dimensions of audience, content, structure and language resources are presented as these are the dimensions that vary according to purpose). The criteria encapsulate the likely development of the dimension so a level 1 performance with regard to structure reads "Some semblance of organization (based around a single idea) may be evident at sentence level and "may attempt conjunctions e.g. and, because etc", while the criteria for level 3 states "Attempts overall structuring of content by grouping ideas within and across sentences; uses simple connectives and linkages within and across sentences e.g. since, though etc, and attempts paragraphing". A group of expert teachers also viewed the criteria, particularly in light of their alignment to curriculum expectations at different levels.

Originally, the rubrics were developed for the first six purposes and for curriculum levels 2 to 4 to cover years 4-8 (Glasswell, Parr & Aikman, 2001). Subsequently, these were extended to level 6 (year 12) and the further purpose, to analyse, was added (Coogan, Hoben & Parr, 2003). For the purposes of both obtaining the national normative sample and then as a resource bank in the subsequently developed tool, appropriate writing tasks related to each purpose were encapsulated in 60 writing prompts that included stimulus material.

The tool was designed to allow teachers to diagnose the writing performance of their students; they are in control; they decide on timing and select and score a task using detailed rubrics, a process that involves moderation and collegial discussion where teachers learn not only about their students but build content knowledge about writing (Parr, Glasswell & Aikman, 2007). The results of the detailed scoring are entered into the software that also contains the normative data. The software is able to generate diagnostic information at individual and group levels. Outputs from the queries, largely presented in visual form, serve a learning function for teachers; they learn about the patterns of strengths and gaps in the writing of their own students both as individuals

and as a group and can view results in relation to various categories of norms. Students can also receive feedback about their individual performance and patterns of relative strengths and weaknesses. The inclusion of the corpus of data from the large normative sample allows teachers not only to consider performance in relation to national average performance for writing for different purposes but also performance may be viewed relative to other factors (e.g. in relation to schools like ours in terms of size and socio-economic standing; my class relative to others of this level; boys in our school relative to all boys; our <ethnic group> students relative to all in that ethnic group etc). Teachers are able to use the information diagnostically, in what Cowie and Bell (1999) term planned formative assessment.

3. Method

3.1 Sample and Procedure

Writing performance data for the data base were obtained from a large, nationally representative sample (N = 20,947) of students in years 4 to 12 of schooling. Sampling was designed to represent proportionally rural and urban schools and types of school. The New Zealand system has both public schools, integrated schools- many of which are catholic- and private schools. In addition, the system has several different configurations in terms of type of school, from area schools in rural communities which cover the range of schooling, to separate primary (years 1-6), intermediate (years 7 & 8) and high schools (years 9-13), plus there are also combinations of these. Schools were then randomly selected. In the administration of the writing tasks to the representative sample of students, a planned missing data matrix design was employed so that an appropriate coverage of writing purposes by year level was ensured. Equal numbers of each writing task were sent to schools and these were systematically assigned by schools to classes (i.e., each class received equal numbers of each task) and the tasks were randomly assigned to students within each class by the teacher. Hence, random assignment of writing tasks and of self-report questionnaires to students was achieved.

Participants produced a writing sample in 40 minutes after brief (five minute) class discussion, designed to stimulate content-related ideas.

3.2 Scoring of Writing

The resulting writing samples were scored using rubrics containing detailed criteria specific to each communicative purpose for writing. The seven different rubrics contained criterion statements at each curriculum level relating to seven dimensions of writing, grouped into two meta-divisions, surface and deep. The seven dimensions or aspects of writing include the four deep features of audience awareness, content or ideas, organisation or structure and language resources and the three more surface features of grammar, spelling and punctuation. While all dimensions of analysis of the text are seen as interdependent in terms of judging the effectiveness of the piece of

writing, for purposes of assessment the dimensions are considered and scored separately.

Markers read each script (the writing produced by an individual student in response to a prompt) and determined which level description best fit each dimension, then decided whether, on each dimension, the score should be augmented (Penny, Johnson, & Gordon, 2000) as being slightly below, or above the assigned level without being in the next level above or below. A weak performance within the level was called Basic, one meeting the criteria in the rubric was called Proficient, and one exceeding the criteria but not yet reaching the next level was called Advanced. Thus each dimension of a script was given a score ranging from Less than Level 2 Basic to Greater than Level 6 Advanced for a total of 17 possible scores.

In the normative sampling, measures were taken to ensure consistency of scoring including training the markers for a half day, having them mark a common script each marking day when a new purpose was scored and then having an expert moderator check their marking every ten scripts or every hour. Daily recalibration exercises were carried out to ensure consistency between days and markers. Marking was deemed to be trustworthy since there was more than 67% adjacent grade consensus, an average Cronbach alpha of greater than .80, and a dependability index (φ) greater than .80 for each purpose scored by one of seven separate panels (Brown, Irving, & Sussex, 2004). Generally, the consistency and reliability of scoring was sufficient to use the data as the basis for national norms and for the data base to serve as a means of enquiring into the performance of students (Brown, Glasswell & Harland, 2004). The Brown, Glasswell and Harland (2004) article also reports a study by Harland that presents a reliability consistency estimate for the four deep feature dimensions and one for the three surface feature dimensions. This provides an indication of whether one was scored more reliably than the other. The consistency estimate for deep features score was $\alpha = .92$ and for surface features it was $\alpha = .93$ suggesting that markers were able to score dimensions like audience or structure as reliably as spelling and grammar.

The total score for each script was found by averaging the seven dimension scores. Curriculum-level scores were transformed to a linear scale; scores were located on a common continuum (Hattie et al, 2003) with a mean of 500 and a standard deviation of 100.

4. Results and Discussion

The normative sample data that comprise the corpus in this instance are able to be interrogated in a number of ways. From a research standpoint, the data can be considered in a cross sectional manner and viewed in terms of patterns of development with age. Performance can be seen in relation to curriculum expectations. The relative performance in each of the different purposes for writing through the course of schooling can be considered. Similarly, the seven different dimensions used to score the writing sample can be considered both by purpose and developmentally across the

year levels of schooling. Or, enquiries can be made regarding differential performance by gender or by ethnicity.

4.1 Development in Writing

Treating the large cross-sectional sample in a longitudinal way allows an enquiry about the relative rate of progress across year levels. With respect to rate of progress over time, the data show that growth, accelerating from year 7, was most marked between years 8 and 10 of schooling (see Figure 1). This takeoff comes, arguably, at a time when opportunity to write increases as writing in secondary school becomes a requirement in most curriculum areas. Writing has been shown to have utility as a tool to support and extend the learning of content in subject areas (Bangert-Drowns, Hurley, & Wilkinson, 2004; Graham & Perin, 2007). Writing is often a catalyst for meaning making and further learning. There is some evidence that a close link to content is significant in enhancing writing performance (Pressley, Mohan, Raphael & Fingeret, 2007). In high school, students may not only have more topic-related knowledge with which to write but also a clearer purpose for writing. The growth curve in writing performance flattens between Years 10 and 12. Arguably, the more specific skill in writing needed to move beyond the level attained by year 10 needs to be both explicitly taught and practised.

Variability in performance across students (in terms of range and standard deviation of total scores) is relatively low at primary but higher at secondary school. This increase in variability around the mean is sometimes referred to as a Matthew effect (Stanovich, 1986), a biblical allusion to St Matthew who wrote of the rich getting richer and the poor getting poorer. It refers to the notion whereby those who perform at a reasonable level in writing build additional skill more readily as they have an existing base of knowledge and skill to build upon. Those who initially have a low level of skill have less to build from. The resulting greater rate of progress of the higher achievers is cumulative and progressively places them further ahead of their peers who were initially lower performing. When the distribution of total writing scores is plotted, there is a large overlap in the distribution of the scores of primary and secondary students; nearly half of secondary students had the same distribution as primary students (Ministry of Education and The University of Auckland, 2006). This widening gap suggests a considerable tail of achievement at secondary level. A plausible, simple explanation may be that lower progress writers simply forgo opportunities to practice the craft. Using data from a study of high and low progress writers (N= 54) in years 1-8 (Glasswell, 1999), the present author calculated how much high progress writers wrote relative to low progress. In the first year of school, when the number of words produced was quite small, high progress writers wrote about twice as many words per sample as struggling writers and there was evidence of this gap widening across the years. For year 1 the effect size for the difference between high progress and struggling writers' word output was 1.01 (Cohen's d) and, by year 8, before entry to secondary school, this difference had risen to 1.59.

4.2 Performance Relative to Curriculum Expectations

As the rubrics for scoring writing are referenced to levels of the national curriculum, it is possible to enquire as to the performance of students, relative to these broad national curriculum expectations. This matching of actual performance to expectations is an informative exercise as curriculum levels are derived from the marrying of theory and practical, professional experience and, therefore, largely represent goals in terms of what is to be aspired to; the skills and knowledge proposed as required at a particular year level. In New Zealand, the curriculum level boundaries regarding what is expected at a particular year level, overlap to take account of student's different individual growth trajectories and the expectation is that students gain a curriculum level every two years. However, matching the closest sub level of a curriculum band expected at a particular year level, shows that students in the representative sample, on average, did not perform well relative to national curriculum expectations. This relatively low level of performance is in line with that reported in the 2007 US National Assessment of Educational Progress where over 50% of Grades 8 and 12 scored at a basic level (Salahu-Din, Persky & Miller, 2008).

It is difficult in the present case to explain such patterns of performance given that little is known about writing classroom practices across the years of schooling, particularly at secondary school level (Kiuhara, Graham & Hawken, 2009). In the New Zealand context, this low performance was not mirrored in reading, so there are clear implications for how writing is taught and teachers supported to teach it. While Kiuhara, Graham and Hawken (2009) note that efforts to improve writing performance are virtually nonexistent in school reform efforts in the United States, this is not the case in New Zealand where 90% of recent schooling improvement has focused on literacy, with equal attention to reading and writing. Further, a school-based, job-embedded national literacy professional development project covering years 1 to 8 of schooling in New Zealand has consistently shown that when teachers are supported to develop appropriate pedagogical content knowledge and to change practice, large gains can be made in writing performance (Parr & Timperley, 2008; Timperley & Parr, 2009). The average effect size gain, over and above expected gain, for three cohorts of schools, calculated using Cohen's *d,* was .76.

4.3 Writing for Different Purposes

Students did not write equally well for all purposes. Although each student in the normative sample only wrote for one communicative purpose, this was a large, representative sample so some conclusions can be drawn about relative competence across purposes. The scores for different writing purposes across year groups are shown in Figure 1. At primary school, mean scores across purposes were relatively similar, on average within 50 points of one another, with explaining and instructing highest and organising and classifying in order to describe and report, the lowest. However, starting at year 9 there is divergence and by years 10-11 there was a considerable gap, up to 150 points, between the highest mean scores (for narrate and report) and the lowest,

recount. It may be that opportunities to practice writing for certain purposes, contract. For example, there may be little call at secondary school to write in order to recount or to instruct. There is support for this in that Kiuhara and colleagues (2009) show that there is a narrowing in terms of the types of writing that teachers report assigning at secondary school. It seems that at year 12 there is some convergence in scores for different purposes, although not to the same extent as in years 4-6. There is still up to 80 points difference between the highest and lowest mean scores, which as at year 10, were for narrate and recount, respectively.

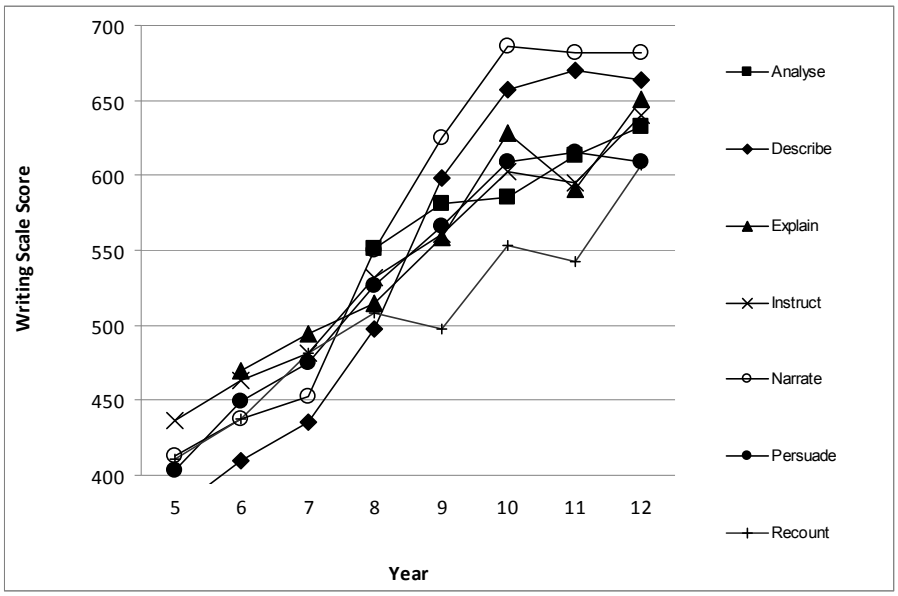

Figure 1. Score by purpose for writing by year of schooling.

This pattern may well be a function of both the complexity in terms of the textual and language demands of the writing purpose and of the emphasis at levels of schooling. There appear to be few studies that analyse the kinds of written text that a particular curriculum area demands in order that students may be supported to enhance their writing skill. Coffin (2006) conducted a discourse analysis of texts and found that the types of writing demanded of the history curriculum in New South Wales were recording, explaining and arguing. Purposes for writing, the types of writing teachers assign and their writing practices vary across subjects at secondary school but, as writing well in several academic areas is a complex skill, the expectation at secondary level might be that there be a shared responsibility across disciplines to teach writing (Graham, 2006). The most frequently occurring writing activities and purposes at secondary school, drawn from a national sample in the United States (Kiuhara, Graham

& Hawken, 2009), involved writing that did not require composing as teachers seldom reported giving essay type assignments. Writing was employed to understand material that had been read and involved activities like summarising and responding. There were, predictably, differences between language arts, science and social studies teachers like the fact that language arts teachers are more likely to use writing for personal expressive or imaginative purposes.

For students in the present data set, at secondary school, reporting is required and practised in subjects like social studies and science while the demands of the subject English in the National Certificate of Educational Achievement undertaken in years 11 and 12 mean that there is an explicit focus on creative storytelling and poetic description. Students represented in the present corpus were less successful with analyse and persuade and, it could be assumed that these are also potential purposes underpinning assigned questions in English and other subjects like history, as Coffin (2006) suggests, at this level. It may be that English teachers, steeped in a literary tradition, are more comfortable with the narrative genre of the novel (the purpose scoring highest) or that the linguistic demands of writing to analyse or to persuade in the context of academic writing, are greater.

While writing may be used in secondary school largely to interact with and learn content, writing serves different purposes in primary school where teachers in New Zealand emphasise the social and communicative aspect. For example, as part of an effort to make the purpose for writing an authentic, communicative one, teachers at primary school in New Zealand use language experience (Parr, Jesson & McNaughton, 2009) so students can practice to narrate, describe, recount, explain and instruct, depending on the nature of the experience. The experience, including the talk that accompanies it, for example, of cooking muffins and following a set of instructions, then later writing instructions for cooking a different type of muffin and perhaps making it, exposes students to the purpose for writing and the text and language features commonly associated with that purpose.

4.4 Performance by Dimension of Writing

The pattern across the seven dimensions of writing showed a similar growth curve to the total score measure. The exception in terms of paralleling the total score growth curve is the area of punctuation where there appears to be a ceiling effect operating by year 8. Perhaps this is partly explained by an inability to demonstrate use of more sophisticated punctuation devices in the course of producing a 40 minute writing sample. Again, there was more convergence across dimensions in mean scores at primary school than later at years 11-12. However, mean scores for the different dimensions were not within measurement error; students were better at some dimensions than others (see Figure 2). This is an interesting finding in light of early discussions concerning holistic versus analytical scoring (e.g. Cooper & Odell, 1977). This, coupled with the different performance across purposes as discussed above, suggests a "one size fits all", analytic rubric may have limitations.

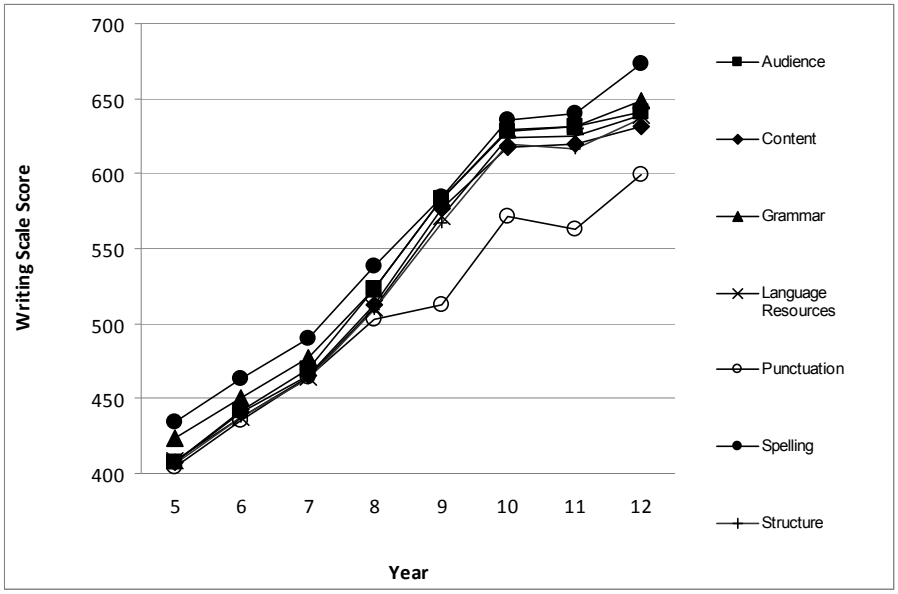

Figure 2. Score for dimensions of writing by year of schooling.

Across all year levels, students were best at spelling and grammar and scored least well on punctuation, followed by structure. Spelling and grammar would seem to be two areas where performance is more clearly able to be specified in an absolute sense. While there was no difference in the reliability consistency estimate for the four deep feature dimensions and that for the three surface feature dimensions of writing, specifying what to teach may be somewhat easier for surface features than for deep features. One of the issues in trying to describe a quality performance in writing concerns the notion that, in writing, what is aimed for is a horizon rather than an end point (Marshall, 2004). In fact, there may be no finality in that it is always possible to craft and improve a piece further, particularly in relation to the deep features.

4.5 Performance and Gender

Writing is perceived as the area where boys significantly underperform in relation to girls; gender differences across subject areas are small, except for writing (Hedges & Nowell, 1995). The perception that girls are more successful than boys at any language-based activity is not new although, as Jones and Myhill (2007) point out, this has not always been seen as problematic. However, a concern that boys are less successful than girls, particularly in writing, is widespread although the extent to which this is actually a new problem is questioned as this gap has been stable for 40 years (Smith, 2003). The significance may lie in the perception of teachers and their assessment bias

towards girls and their under rating of boys' performance (Reeves, Boyle & Christie, 2001).

The analysis of this corpus of data showed that girls performed significantly better across all year groups (average effect size .43, Ministry of Education and The University of Auckland, 2006), with the gap increasing after year 7 (Grade 6) through secondary school to be at its greatest extent at year 9 (Grade 8) where girls scored on average 80 points or two school years ahead of boys. This trend is similar to what is generally reported internationally except that, in the current data, the gap narrows by years 11 and 12 of schooling. This is at variance with the analysis of the National Association for Educational Progress writing data in the US for 2002 where the gap is greater at Grade 12 than Grade 8 (Smith & Wilhelm, 2009).

There have been suggestions that the way writing achievement is measured may underestimate the performance of boys; there is evidence that girls' narrative writing may be privileged in assessments (Peterson, 1999) and their writing more aligned with the approved literary canon than boys. Boys are not deficient in writing, they are differentially literate (Millard, 1997). It has been suggested that boys reject literacy forms that are schoolish and divorced from students' home literacies (Cavazos-Kottke, 2005). Although Daly (2002) argues that we over generalise the belief that boys are more disposed to non-fiction texts, in the present data, there is some support for a narrowing of the gender gap in performance according to the purpose for writing. This is contrary to the results of Jones and Myhill (2007) who found that text type (in their case) had no impact on the level of difference. When writing score in the present data is considered according to purpose for writing, by year, for males and females separately, the confidence intervals for explain, analyse, recount, instruct and persuade overlap. Bar recount, these are purposes for writing that largely involve the transaction of information. Consistent with the data for overall average performance, where the gap narrowed at years 11 and 12, most of the overlaps are at these year levels.

As would be expected from the differences in total score, males and females performed differently, on average, on both deep and surface features of writing; dimensions with overlapping confidence intervals at year level were audience and grammar (Yr 11).

There is support for the argument that, while boys may not be in crisis as there is strong evidence of academic improvement- it is simply that girls have made greater improvement-, there are groups of boys at risk (Mead, 2006). In these data there is evidence that boys from particular minority ethnic groups may be more at risk.

5. Conclusion

The corpus that was the subject of this paper is able to serve dual purposes- research and pedagogical. Here, the normative sample data underpinning the diagnostic pedagogical tool have been questioned with a research focus. Questioning an extensive mediated corpus of writing data yields interesting results regarding patterns of

performance across age with respect to both purposes for, and dimensions of, writing. It allows an examination of performance by gender.

But, the use of such a corpus also resurfaces issues with respect to the mediating process by which the original handwritten scripts were transformed into a series of scores. These issues concern decisions regarding defining the different communicative purposes for writing; the operationalising of the dimensions of writing through specifying the criteria to be scored and the consistency with which this can be done. The research literature concerning communicative purposes for writing, together with the national curriculum informed the delineation of the different writing purposes. The dimensions of writing, similarly, drew on research in language and in writing. The more problematic element of operationalising the aspects of writing, the construction of the detailed criteria for scoring each of the dimensions within each of the writing purposes, was multiply informed. The work of linguists and of genre theorists was synthesised by the researchers. The resulting criteria were then examined in the light of the professional knowledge of writing practitioners. It seems that in order to understand progression in writing, teachers may rely less on a knowledge model specifying precisely what makes for a quality performance than on value judgements that are formed through the extensive process of making those judgements. In this way, expectations of performance at different points in development are established. Sadler wrote of the guild knowledge of teachers. The suggestion is that teachers come to make reliable judgements of written work through the process of construct referencing. The understanding of a construct is refined through experience and processes like moderation where collegial discussion is involved. The shared meaning that develops among those interpreting the evidence in the form of the writing within a context is the guild knowledge (Sadler, 1987, 1989). It is this knowledge that was married with the work of the linguists and genre theorists to construct the final scoring rubrics that were used to mediate the handwritten responses of students to the specially constructed prompts. The samples were able to be scored consistently and the resulting scores used to build the data base, the corpus of student writing performance to interrogate.

The results of this interrogation show that progress in writing accelerates around year 8 of schooling (Grade 7) although the explanation for this is somewhat speculative as there are no related systematic data on teacher practice, unlike the data reported recently by Kiuhara, Graham and Hawken (2009), in the US. Without confirmation from such data about the influence of teaching practices we cannot discount competing hypotheses for the acceleration like the notion that around year 8 students start to read more complex and diverse texts and this provides them with ideas and potential models. It is also at this point that writing is employed more in the service of learning and it may be more a function of opportunity to practice than a direct result of teaching practice. As there is a lack of evidence that teachers in different subject areas actually teach the features of the writing demanded in their subject or discipline area, opportunity to practice is the most likely explanation for the finding that performance across purposes varies and at different points in development, the relative difference in

performance by purpose is greater. This is likely to be a function of the demands of different school subject choices and of the nature of writing associated with them.

The fact that with respect to dimensions of writing students performed best on surface features of spelling and grammar appears not to be a function of more consistent scoring because of the potential for more specifically defined criteria but may relate rather to the concrete nature of such features helping teachers to articulate the goals more readily.

Students in this normative sample did not meet curriculum expectations in writing; although curriculum expectations tend to represent aspirations as opposed to what students can actually do. This finding, together with the allied finding that students are two years behind curriculum expectations in writing by the end of school but that they meet curriculum expectations in reading and math has led to targeted efforts to raise achievement through national literacy professional development and schooling improvement efforts focussing on literacy.

The data regarding performance by boys and girls reinforce findings from the literature showing that their performance differs markedly. However, the closing of the gap towards the end of school is interesting to view in the light of social and cultural factors that may be operating at this level as students prepare for tertiary study or the workplace. Writing may appear more purposeful and less schoolish for boys and the fact that their scores approach those of girls in transactional forms of writing lends some support to this notion.

These differences in patterns of growth across years of schooling, including the pattern by gender and that in regard to writing for different purposes underscore the need for practitioners to be aware of and to attend to diversity and difference. The other purpose of this dual purpose data base is to allow teachers to do just this- to attend to diversity and difference within their writing classrooms by utilising the powerful diagnostic tool that has been built from this corpora of writing data.

Note

1. The tool is part of Project asTTle (Assessment Tools for Teaching and Learning) which provides detailed assessment against curriculum objectives in reading, writing and mathematics for Years 4 to 12 (a full description of this project along with technical reports and publications is available at http://www.tki.org.nz/r/asttle). It is a CD Rom-based/ web based assessment suite which gives teachers choice in the design and timing of assessments and access to a range of reporting formats, including comparisons to norms.

References

Bangert-Drowns, R.L., Hurley, M.M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74*, 29-58. doi: 10.3102/00346543074001029

Biber, D. (1993). Using register-diversified corpora for general language studies. *Computational Linguistics*, *19*, 219-241.

Braun, S. (2007). Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL, 19*, 307-328. doi: 10.1017/S0958344007000535

Braun, S., Kohn, K., & Mukherjee, J. (Eds.). (2006). *Corpus technology and language pedagogy: New resources, new tools, new methods*. Frankfurt, Germany: Peter Lang.

Brown, G.T.L., Glasswell, K., & Harland, D. (2004). Accuracy in scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing*, *9*, 105-121. doi: 10.1016/j.asw.2004.07.001

Brown, G.T.L., Irving, S.E. & Sussex, K. (2004). *Scoring of a nationally representative sample of student writing at level 5—6 of the New Zealand English curriculum: Use and refinement of the asTTle progress indicators and tasks* (Project asTTle Tech. Rep. No. 48). Auckland, New Zealand: University of Auckland. Retrieved from http://www.tki.org.nz/r/asttle/pdf/technical-reports/techreport48.pdf

Cavazos-Kottke, S. (2005). Tuned out but turned on: Boys (dis)engaged reading in and out of school. *Journal of Adolescent and Adult Literacy*, 49, 180-184. doi: 10.1598/JAAL.49.3.1

Coffin, C. (2006). Learning the language of school history: The role of linguistics in mapping the writing demands of the secondary school curriculum. *Journal of Curriculum Studies*, *38*, 413-429. doi: 10.1080/00220270500508810

Coogan, P., Hoben, N., & Parr, J. (2003). *Written language curriculum framework and map: Levels 5-6*. (Project asTTle Tech. Rep. No. 37). Auckland, New Zealand: University of Auckland. Retrieved from http://www.tki.org.nz/r/asttle/pdf/technical-reports/techreport37.pdf

Cooper, C.R., & Odell, L. (1977). *Evaluating writing: describing, measuring, judging*. Urbana, IL: National Council of Teachers

Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education, 6,* 101-116. doi: 10.1080/09695949993026

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*, 213-238. doi: 10.2307/3587951

Daly, C. (2002). *Literature search on improving boy's writing*. London, England: Office for Standards in Education.

Derewianka, B. (1990). *Exploring how texts work*. Sydney, Australia: Primary Teachers Association.

Glasswell, K. (1999). *The patterning of difference: Teachers and children constructing development in writing* (Unpublished doctoral thesis). University of Auckland, Auckland, New Zealand.

Glasswell, K, Parr, J.M., & Aikman, M. (2001). *Development of the asTTle writing assessment rubrics for scoring extended writing tasks* (Project asTTle Tech. Rep. No. 6). Auckland, New Zealand: University of Auckland. Retrieved from http://www.tki.org.nz/r/asttle/pdf/technical-reports/techreport06.pdf

Graham, S. (2006). Writing. In P. Alexander & P. Winne (Eds.), *Handbook of Educational Psychology* (pp. 457-477). Mahwah, NJ: Erlbaum.

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, *99,* 445-476. doi: 10.1037/0022-0663.99.3.445

Hattie, J.A.C., Brown, G.T.L., Keegan, P.J., Irving, S.E., MacKay, A.J., Pipi, A.H., . . . Mooyman, D. (2003). *Assessment Tools for Teaching and Learning (asTTle) Manual: Version 3 2004*. Wellington, New Zealand: Learning Media.

Hayes, J. R., Hatch, J. A., & Silk, C.M. (2000). Does holistic assessment predict writing performance? Estimating the consistency of student performance on holistically scored writing assessments. *Written Communication, 17*, 3-26. doi: 10.1177/0741088300017001001

Hedges, L.V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high scoring individuals. *Science*, *269*, 41-45. doi: 10.1126/science.7604277

Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *Language Learning Journal*, *36*, 153-166. doi: 10.1080/09571730802389983

Hyland, K. (2006). *English for academic purposes: An advanced resource book*. London, England: Routledge.

Jones, S., & Myhill, D. (2007). Discourses of difference? Examining gender differences in linguistic characteristics of writing. *Canadian Journal of Education*, *30*, 456-482. doi: 10.2307/20466646

Kiuhara, S.A., Graham, S., & Hawken, L.S. (2009). Teaching writing to high school students: A national survey. *Journal of Educational Psychology*, *101*, 136-160. doi: 10.1037/a0013097

Knapp, P., & Watkins, M. (1994). *Context, text, grammar: Teaching the genres and grammar of school writing in infant and primary classrooms*. Sydney, Australia: Text Productions.

Knapp, P., & Watkins, M. (2005). *Genre, text, grammar: Technologies for teaching and assessing writing*. Sydney, Australia: University of New South Wales Press.

Kress, G. (1993). Genre as social process. In B. Cope & M. Kalantzis (Eds.), *The powers of literacy: A genre approach to teaching writing.* (pp.22-37). Pittsburg, PA: University of Pittsburg Press.

Krishnamurty, R., & Kosem, I. (2007). Issues in creating a corpus for EAP pedagogy and research. *Journal of English for Academic Purposes*, *6*, 356-373. doi: 10.1016/j.jeap.2007.09.003

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*, 307-322. doi: 10.1093/applin/16.3.307

McCarthy, M. (2001). *Issues in applied linguistics*. Cambridge, England: Cambridge University Press.

Marshall, B. (2004). Goals or horizons- the conundrum of progression in English: Or a possible way of understanding formative assessment in English. *The Curriculum Journal*, *15*, 101-113. doi: 10.1080/0958517042000226784

Martin, J.R., Christie, F., & Rothery, J. (1987). Social processes in education. In I. Reid (Ed.), *The place of genre in learning: Current debates*. Geelong, Australia: Deakin University Press.

Mead, S. (2006). *The evidence suggests otherwise: The truth about boys and girls*. Washington, DC: Education Sector. Retrieved from http://www.educationsector.org/usr_doc/ESO_BoysandGirls.pdf

Merlo, P. (1994). A corpus-based analysis of verb continuation frequencies for syntactic processing. *Journal of Psycholinguistic Research*, *23*, 435-447. doi: 10.1007/BF02146684

Millard, E. (1997). *Differently literate*. New York, NY: Falmer.

Ministry of Education. (2006). *Student achievement in New Zealand*. Wellington, New Zealand: Author.

Nesi, H., Gardner, S., Forsyth, R., Hindle, D., Wickens, P., Ebeling, S., . . . Heuboeck, A. (2005). Towards the compilation of a corpus of assessed student writing: An account of work in progress. *Proceedings from the Corpus Linguistics Conference Series Volume 1*. Retrieved from http://www.corpus.bham.ac.uk/pclc/index.shtm1

Parr, J.M., Glasswell, K., & Aikman, M. (2007). Supporting teacher learning and informed practice in writing through assessment tools for teaching and learning. *Asia-Pacific Journal of Teacher Education*, *35*, 69-87. doi: 10.1080/13598660601111281

Parr, J.M., Jesson, R., & McNaughton, S. (2009). Agency and platform: The relationships between talk and writing. In J. Riley, R. Beard, D. Myhill & M. Nystrand (Eds.), *Handbook of Writing Development* (pp. 246-259. London, England: Sage. doi: 10.4135/9780857021069.n17

Parr, J.M., & Timperley, H. (2008, April). *Assessment for teaching and learning in writing classrooms: Building teacher knowledge enhances student achievement.* Paper presented at the Annual Meeting of the American Educational Research Association, New York.

Penny, J., Johnson, R. L., & Gordon, B. (2000). The effect of rating augmentation on inter-rater reliability: An empirical study of a holistic rubric. *Assessing Writing, 7*, 143-164. doi: 10.1016/S1075-2935(00)00012-X

Peterson, S. (1998). Evaluation and teachers' perception of gender in sixth grade student writing. *Research in the Teaching of English*, *33*, 181-206.

Pressley, M., Mohan, L., Raphael, L. M., & Fingeret, L. (2007). How does Bennett Woods Elementary School produce such high reading and writing achievement? *Journal of Educational Psychology, 99*, 221-240. doi: 10.1037/0022-0663.99.2.221

Reeves, D.J., Boyle, W.F., & Christie, T. (2001). The relationship between teacher assessment and pupil attainments in standard test/tasks at Key Stage 2, 1996-8. *British Educational Research Journal*, *27*, 141-160. doi: 10.1080/01411920120037108

Roland, D., Dick, F., & Elman, J.L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, *57*, 348-379. doi: 10.1016/j.jml.2007.03.002

Sadler, D.R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education, 13*, 191-209. doi: 10.1080/0305498870130207

Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*, 119-44. doi: 10.1007/BF00117714

Salahu-Din, D., Persky, H., & Miller, J. (2008). *The nation's report card: Writing 2007* (Report No. NCES 2008-468). Washington, DC: US Department of Education. Retrieved from http://nces.ed.gov/nationsreportcard/pubs/main2007/2008468.asp

Smith, E. (2003). Understanding underachievement: An investigation into the differential attainment of secondary school pupils. *British Journal of Sociology of Education*, *24*, 575-586. doi: 10.1080/0142569032000127143

Smith, M.W., & Wilhelm, J.D. (2009). Boys and literacy. In L. Christenbury, R. Bomar & P. Smagorinsky (Eds.), *Handbook of adolescent literacy research* (pp. 360-371). New York, NY: Guilford Press.

Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360-401. doi: 10.1598/RRQ.21.4.1

Timperley, H., & Parr, J.M. (2009). What is this lesson about?: Instructional processes and student understandings in the writing classroom. *The Curriculum Journal, 20,* 43-60.doi: 10.1080/09585170902763999

Vannestal, M., & Lindquist, H. (2007). Learning English grammar with a corpus: Experimenting with concordancing in a university grammar course. *ReCALL*, *19*, 329-350. doi: 10.1017/S0958344007000638

Williamson, M., & Huot, B. (1993). *Validating holistic scoring for writing assessment: Theoretical and empirical foundations*. Cresskill, NJ: Hampton.

Wing Jan, L. (1991). *Write ways: Modeling writing forms*. Melbourne, Australia: Oxford University Press.

Appendix A

Scoring Rubric, Purpose: PERSUADE

	Level 1 (Proficient)	Level 2 (Proficient)	Level 3 (Proficient)	Level 4 (Proficient)	Level 5 (Proficient)	Level 6 (Proficient)
Audience Awareness and Purpose	Writer writes primarily for self.	Writer recognizes they are writing for an audience other than self.	Shows some awareness of purpose and audience through choice of content, language and writing style.	Writer shows awareness of purpose and audience through choice of content, language and writing style.	Writer shows awareness of purpose and targets the audience through deliberate choice of content, language and writing style.	Writer consistently persuades intended audience.
	States own opinion with little attempt to persuade.	May attempt to persuade audience.	Attempts to persuade the audience by stating position in opening.	Clearly states a consistent position to persuade the audience.	Identifies and relates to a concrete/specific audience.	Shows implicit awareness that audience may hold a range of points of view.
	States opinions from a personal perspective and assumes shared knowledge with the audience.	States opinions from a personal perspective and may assume shared knowledge with the audience.	Knows that audience may hold a different point of view but tends to assume there is only one generalized point of view.	Shows some awareness of intended audience particularly at beginning and end of text.	Shows awareness of intended audience and acknowledges others' point of view.	Uses tone for impact or to manipulate the intended audience towards author's point of view. May effect change.

Content/Ideas	Writing includes one or more domains appropriate to purpose, usually a position statement that conveys a simple idea or a response from a personal perspective.	Writing includes some domains appropriate to purpose, e.g. a position statement in which the writer identifies a position and makes two or more simple related opinions or statements.	Includes most domain elements for argument, e.g. main points, some supporting evidence or illustration, a re-statement of position.	Includes and begins to develop identifiably domain elements for argument, e.g. a position statement, support for main points, restatement.	Develops mainly consistent domain elements for argument, e.g. a plausible position statement, support for main points, restatement.	Selects content to add. Makes considered, relevant and elaborate points. Chooses examples to support purpose.
		May include a conclusion.	May include a conclusion that makes a recom-mendation.	Restates and strengthens position.	Uses conclusion to reflect points made and may expand the argument.	Uses conclusion to integrate the themes of the argument, rather than simply repeating or summarising the points made.
	May repeat some ideas.	May present ideas as a list.				
	May include information unrelated to the topic and/or task.	May include some statements unrelated to the topic and/or task.	Relates almost all material to the given task.	Provides relevant support for ideas.	Strongly links supporting reasons to argument.	Gives consistent support to main points.

Structure	Some semblance of organisation (based around a single idea) may be evident at sentence level.	Semblance of organisation e.g., some grouping of ideas, generally at sentence level, is evident. May make opinion statements as discrete elements.	Attempts overall structuring of content by grouping ideas within and across sentences.	Groups content logically at the level of main idea by using topic sentences to guide the reader's understanding.	Uses structure to add to the intended impact of argument e.g., by developing a logical, consistently flowing argument.	Uses an explicit, logical structure to enhance the argument.
	May attempt simple conjunctions e.g., "and", "because", etc.	Attempts simple conjunctions to link ideas within sentences, e.g., "and", "because", etc.	Uses simple connectives and linkages within and across sentences, e.g., "since", "though", etc.	Consistently uses a variety of connectives and linkages within sentences and between paragraphs, e.g., "on the one hand", "however", etc.	Uses complex linkages within and between paragraphs, e.g., varied linking words and phrases, conjunctions and text connectives.	Uses complex linkages, e.g., varied linking words and phrases, conjunctions and text connectives.
			Attempts paragraphing.	Uses paragraphing, linking main ideas and supporting details.	Uses paragraphs with main ideas and supporting details. Links sentences thematically to topic or paragraph or section.	Uses logically arranged reasoned ideas in well-crafted paragraphs and strong topic sentences to guide the readers understanding of the argument.
Language Resources	Uses simple opinion statements from a personal perspective, e.g., "I like", etc.	Uses simple persuasive statements from a personal perspective, e.g., "I think", etc.	Uses some features of persuasive language e.g. rhetorical questions, imperatives, passive voice, data.	Uses features of persuasive language, e.g., rhetorical questions, imperatives, passive voice, data.	Deliberately uses a range of features of persuasive language for effect in order to involve and persuade the intended audience.	Uses language features for effect to involve and persuade the intended audience.

Uses some topic-specific language to express an opinion. Uses mainly high-frequency words.	Uses topic or content-specific language but language choices convey little opinion, e.g., mainly neutral nouns, basic descriptors and limited verbs and adverbials.	Begins to select language to create a particular effect to influence the audience, e.g., 'point of view' nouns, viewpoint adverbials and opinion adjectives to add detail and weight to opinion statements and evidence. May use some modal auxiliary verbs, e.g., "can", "might", "should", "may", etc.	Uses language to identify a particular viewpoint and persuade the audience.	Uses passive structure and modal auxiliaries to strengthen argument.	Considers and selects language features for effect with the intention of manipulating and/or influencing the audience. Uses tone, e.g., sarcastic, threatening, humourous, emotive etc, to underpin selective language features and strengthen argument.
Shows some understanding of pronoun use.	Shows some understanding of pronoun use.	Largely controls pronoun use.			