Oops, that was a mistake: Examining the effects and implications of changing assessment policies.

Paper presented at the Empirical Issues in Canadian Education Conference
November 23-24, 2001

Don Klinger
klingerd@educ.queensu.ca
A-206 Duncan McArthur Hall
Faculty of Education
Queen's University
Kingston, Ontario
K7L 3N6

Oops, that was a mistake: Examining the effects and implications of changing assessment policies.

Introduction

The use of large-scale testing is varied and expanding in terms of purpose, function, and analysis. In Canada, British Columbia, Alberta, Manitoba, Quebec, New Brunswick, and Newfoundland have provincial examination programs that are used to help determine High-School students' grades (Cheliminsky & York, 1994; Lafleur & Ireland, 1999). Ontario is introducing a grade 10 literacy examination that students will be required to pass in order to graduate and a teacher certification examination that new teacher graduates must successfully complete before becoming qualified to teach in Ontario. Other large scale assessments are in place or being implemented to support accreditation, monitor student achievement, or measure change. As these programs are introduced or fully implemented, it is not unusual for changes or modifications to occur to the assessment program. While some modifications could be considered part of the maturation process of the program, others are the result of financial constraints, changing purposes, or new methodologies.

Policy makers and stakeholders are generally interested in the results associated with assessment programs. Researchers examining such programs pay particular attention to issues of psychometrics, equity, technology, and form (e.g., Kolen & Brennan, 1996; Muraki, 1992; Shealy & Stout, 1993; Yen & Ferrara, 1997). As suggested by Delandshere (2001) "Current assessment practices are debated with regard to the impact they have on defining or narrowing the curriculum, for example, or on the need for different forms of assessment (e.g., Shepard, 1989)." (p. 115). Developing technologies, conceptual and theoretical advancements, financial constraints, and curricular shifts

ensure that changes and modifications to large scale assessment programs are, and will continue to be, inevitable (Bennett, 1998). Given this environment of change, comparability of results is an important consideration. Nonetheless, format or procedural changes should not generally result in different decisions or results (Thomasson, 1997). Unfortunately, variations that arise due to changes in format, procedure, reliability, scoring methods, or standard setting can threaten comparability and utility (e.g. Livingston & Zieky, 1983; Plake & Giraud, 1998; Wainer & Thissen, 1993). Further, it is often difficult to anticipate the effects and implications associated with such modifications on those involved with or interested in the results of such assessments.

As an example, the choice of item format has been shown to have differential effects on subgroups of the population. Research suggests that males do better on multiple-choice (MC) items while females do better on extended-response (ER) items (e.g., Bolger & Kellaghan, 1990, Garner & Engelhard, 1999; Henderson, 1999). These gender effects are even more pronounced for high achieving students (Bridgeman, 1989; DeMars, 1998; Schmitt, Mazzeo, and Bleistein, 1991). Based on these findings, the addition of ER items would likely benefit female students while the addition of MC items would more likely benefit male students. Hence, format changes that affect the ratio of MC and ER items on an assessment may be harmful or beneficial to different groups of examinees.

Research into large-scale assessment programs has focussed largely on issues of fairness, equity, and form, with a heavy emphasis on “technique” and other technical issues. In practice, these issues often receive much less attention, if at all. While the results of assessments are used to debate the quality of schooling, the assessment

practices upon which these debates are based remain largely unquestioned (Delandshere, 2001). There seems to be an implicit assumption made by examinees, teachers, administrators, and policy makers that these issues are not important in relation to the conclusions to be made nor will changes in format and procedure affect the results and findings of assessments. Consequently, there has been little research examining the effects of such changes with respect to their influence on policy decisions and directions. High stakes assessment programs have strong consequences and research is needed in Canada as well as elsewhere to evaluate the degree to which these assessments meet and continue to meet expectations (Baker, O'Neil, & Linn, 1994). Specifically, given the increased uses and consequences attached to assessment results and given the pace at which such assessments are being implemented and modified, it is important to examine the effects of changing assessment procedures both from a technical and policy perspective. Using examples drawn from two different sources, the first from research and the second from professional observation, the current study examines the implications and consequences of changes in assessment programs as they impact students, teachers, and decision makers. Based on these findings and observations, guidelines and recommendations are provided for future large-scale assessment programs.

The British Columbia scholarship program.

Students in British Columbia High-School complete curriculum based examinations in each of their grade 12 academic courses, with the results providing 40% of their final course mark. However, a unique aspect of the examination program is that students can also obtain provincial scholarships based on examination performance. In

the years between 1984 and 1996, students interested in obtaining a provincial scholarship not only wrote the mandatory two hour provincial examinations in the academic grade 12 courses in which they were enrolled but also wrote the optional one hour scholarship examinations for these courses. Each of the provincial examinations consisted of a set of multiple-choice (MC) and extended-response (ER) items that encompassed the major concepts within each curriculum. In contrast, the scholarship examinations consisted only of ER items that were considered to be conceptually more difficult. The scholarship examinations had no impact on students' marks but were used in conjunction with the corresponding provincial examinations to determine subject level scholarship scores. A student's raw scholarship score was calculated by adding the provincial and scholarship examination scores. Since the scholarship examination was only one hour in length, the contribution of each scholarship examination was approximately one-third while the provincial examination contributed two-thirds towards the scholarship score. The actual contribution of each component of the examination would vary depending on the standard deviation of the components. For each subject, the set of scholarship scores were normalized using the RANKIT procedure (Chambers, Cleveland, Kleiner, & Tukey, 1983) and then transformed onto a standardized scale having a mean of 500 and a standard deviation of 100, with minimum and maximum scores set to 200 and 800, respectively. Adjustments were made to the highest scores to ensure that the maximum scholarship score in each subject was in fact 800. Scholarships were awarded to students obtaining a scholarship score of at least 475 based on their three highest examinations and having a combined minimum total of 1700 using these three scores.

Beginning with the 1996/97 school year, the provincial government introduced the current procedure, eliminating the scholarship examinations but maintaining the scholarship program itself. Thus, with the current procedure, scholarships are still awarded to students based on the formulas above, but the calculation of scholarship scores and scholarship decisions are now based solely on the provincial examinations. The content, construction, and format of the provincial examinations have not changed. However, to mimic the sample of students who typically wrote scholarship examinations, only those students having a provincial examination score of 70% or higher receive a scholarship score for that examination.

One reason for the elimination of the scholarship examinations was the cost of development and marking of these examinations during a period of time when government expenditures were being reduced and departments were asked to find ways to reduce costs. The belief was that the information gained from the scholarship examinations was redundant with the information from the provincial examinations because the scores from both examinations were highly correlated (Ron West, personal communication, September 22, 1999). Nonetheless, since there has not been a change in the difficulty of the provincial examinations, the examinations used to determine scholarship recipients are now shorter and simpler. This, in turn, may result in different students receiving scholarships than would have if the scholarship examinations had not been removed.

To address the effects of the examination format changes, analysis was based on the last two school years in which the one-hour scholarship examinations were written, 1994/1995 and 1995/1996. The actual scholarship scores and decisions obtained by

students, based on both the provincial and scholarship examinations, were considered the 'gold standard.' New scholarship scores and decisions were then generated using the current procedure (provincial examination scores) and compared to the gold standard. Results were based on the January and June versions of the grade 12 Biology, Chemistry, Français Langue, French, Geography, Geology, History, English Literature, Mathematics, Mandarin, and Physics. English 12 examination results were not included because there were no changes to the method in which scholarship scores were calculated for the English examination. The 1995 June French 12 examination was not included in the analysis because of a security breach, negating the use of the scholarship examination. November, April, and August examinations as well as the other language examinations (Latin, German, Spanish) were not included because of very small sample sizes. Finally, given the optional nature of the scholarship examinations, comparisons were based only on those students who chose to write the scholarship examinations.

Comparisons between the current and original procedure (the gold standard) were completed at both the subject and total score level. At the subject score level, correlations, root mean square error analysis (RMSE), and decision consistency (scores above or below 475) were examined. At the total score level, the identification of scholarship recipients and non recipients was compared between the two procedures. Further analyses were completed to determine if any differences could be attributed to systematic effects.

Differences between the current and original procedure exist at the subject level and these differences vary across subjects. The median correlation between current and original procedures was 0.87 with a low of 0.70 for Français langue, June 1995, and a

high of 0.92 for Biology June 1996. RMSE values illustrate similar variation. As summarized in Table 1, RMSE values varied from a value of 38.7 (June 1996 Biology) to 80.4 (January 1995 Literature). With respect to decision consistency, the examination classification error rates varied from a low of 8.9% (June 1995 Biology) to a high of 26.0% (June 1995 Français Langue). Increased errors were associated with examination periods and specific examinations. With the exception of the 1995/1996 French and Physics examinations, higher RMSE values and classification errors were found for the January examinations than for the corresponding June examinations. The majority of the errors were false negative classifications in which a student would not obtain the minimum scholarship score of at least 475 if the current procedure was used although they actually did obtain the minimum score. The Mandarin examination was the one exception to this pattern, having higher false positive rates. False negative classifications were three times more likely in the January examinations as compared to the corresponding June examinations. Thus, at the examination level, the current procedure generally underestimated the number of students who would receive the minimum scholarship score and this problem was more pronounced during the January examinations.

Insert Table 1 about here

Other systematic errors were also present across the set of examinations. With a negative correlation of -0.63 ($p<0.01$), smaller sample size was associated with higher error rates using the current procedure. Increased errors were also associated with specific examinations. Those examinations having a subjective scoring procedure (Français Langue, Geography, History, and Literature) tended to have the highest RMSE

values and error rates, while those examinations having more defined analytical scoring procedures (e.g., Biology, Chemistry, Mathematics, Physics) had lower values and rates. Interestingly, false negative and false positive classifications were similar for both males and females, indicating that the removal of the ER items did not differentially impact males or females.

The last row of Table 1 summarizes the errors associated with the awarding of scholarships. The use of the three highest scholarship scores above 475 and having a combined score of at least 1700 moderated the error rates observed at the examination level. Nonetheless, at 8.2% (7.7% in 1994/1995 and 8.4% in 1995/1996) these error rates remained substantial. With a ratio of 1.6 to 1, the majority of errors represented false negative decisions, in which students would have been unfairly denied a scholarship if the current procedure had been in place.

The British Columbia Foundation Skills Assessment

Beginning with the 1998/1999 school year, the British Columbia Ministry of Education introduced the Foundation Skills Assessment (FSA) program. All grade 4, 7, and 10 students, with few exceptions, are expected to write separate assessments in numeracy, reading comprehension, and writing. The assessments are developed and scored by teams of teachers. Based on their performance, students are placed in one of three categories for each of the three domains; 1) not yet within expectations, 2) within expectations, and 3) exceeds expectations. The assessments do not affect students' grades in any way; however, the results are used to provide an external measure of individual student achievement, measure overall achievement at each grade, and provide data to support accreditation of schools. Implementation of the FSA has been characterized by

change. While many of the changes implemented by the Ministry of Education have been in response to concerns raised by school and district personnel. However, these modifications have also created problems.

During the first year of implementation, the Ministry of Education retracted its decision to provide student or school level information. While the decision was heralded as a victory by the British Columbia Teachers Federation, it also created problems for several schools that were intending on using the data to support school growth plans and accreditation. For example, Bulkley Valley, a small district in rural north western British Columbia, was forced to drop plans to use the FSA writing assessment as an external measure of students' writing skills. With increased writing achievement as a district wide growth target, a district based assessment and scoring procedure was developed and used as the sole external measure for determining a district baseline and implementing a monitoring process. Interestingly, student and school level information was provided during the second year (1999/2000) of the FSA program, albeit in a format unfamiliar to students, teachers, district personnel, or parents. After the second year of implementation, further concerns were raised regarding the writing component of the FSA. In particular, the use of the same prompt at all levels and the use of a single writing sample to determine writing proficiency. In voicing these concerns, teachers were questioning the results. The writing assessment was changed during the third year (2000/2001) of the program, using two samples on which to base decisions and having grade specific writing prompts. The evaluation system was also modified to incorporate Performance Standards that had been developed and modified by educational leaders and teachers. With the

change in the assessment and marking system, new standards were created in August of 2001.

Once again, these changes were largely in response to concerns raised by teachers and evaluation specialists. However, in implementing these changes, the previous results and efforts were essentially negated. As an example, the Assistant Superintendent of Bulkley Valley recently asked the following questions: "Since the Ministry of Education has now adopted the performance standards for the scoring of the FSA writing assessments, how can we move to incorporate the performance standards into our district wide assessments? What will the switch do to our baseline data? Will we be starting all over again? Will this mean we have to set standards all over again?" (Judy Morgan, personal communication, October 30, 2001). At the provincial level, the change in procedure has also led to very different findings regarding the level of writing for grade 10 students in British Columbia. A review of the 2000 results indicates that 32% of grade 10 students were categorized as "not yet within expectations" whereas the 2001 results placed 14% of students in this category (British Columbia Ministry of Education, 2000; 2001).

Discussion

Modifications and changes to assessment programs have implications for students, teachers, and policy makers. In examining new or changing assessment procedures, policy makers tend to look for overall similarities between procedures. This often cursory comparison may in fact mask important and unexpected differences in the results. The removal of the scholarship examinations was based largely on high correlations, a belief that now appears misguided. The current procedure has resulted in

increased errors and a majority of these errors unfairly deny students a provincial scholarship. Further, the removal of the scholarship examinations has had differential effects across the various examinations. This would make some courses more or less desirable for students interested in obtaining scholarships. Fortunately, In spite of previous research suggesting differential results associated with gender, the removal of the scholarship examinations has affected both males and females equally. Finally, the removal of the scholarship examinations may have also changed teaching practices and modified the curriculum being taught in grade 12 academic courses. Since the provincial examinations have not changed, some of the more difficult concepts are no longer being examined. When the scholarship examinations were in place, teachers would spend class or extra-curricular time, to prepare students for the scholarship examinations. This incentive is no longer present. Further research is required to examine the impact the removal of the scholarship examinations has had on teacher practice

In the case of the Foundation Skills Assessment, the changing policy first in terms of the method of reporting results, second in the addition of a second writing sample, and third in the change in scoring has, at the very least, served to delay the implementation of school and district based improvement plans that were to use these results as indicators. Further, those districts and schools that did the most work to support the use of the assessments were the ones most negatively affected by the delays and changes. For example, Bulkley Valley's three year writing assessment will require additional expenditures if the district wishes to unify district and provincial procedures.

However, the changes have also undermined the previous FSA results, the assessments as a whole, and may lead to misinformed policy decisions. The dramatic

drop in the number of students not yet within expectations is problematic. In 2000, the writing results were cause for alarm in secondary schools across the province. In contrast, the 2001 results place writing in a much lower priority than either reading or numeracy, with respectively, 25% and 23% not yet within expectations. It is unlikely this "effect" is due to changes in educational practice, especially since such changes were not found at the other grade levels. Rather, the changes are likely associated with the change in marking procedures and the subsequent standard setting. Unfortunately, given the lack of assessment expertise of most teachers, administrators, district personnel and the public at large, inappropriate comparisons will be made between the two years and the differences attributed to a variety of factors, both real and imagined. In particular, the drop in the number of students not yet within expectations may be viewed as evidence that recent educational policies have been effective in addressing writing concerns. Nor has the Ministry of Education done enough to prevent inappropriate comparisons. While the Ministry report states that new standards were developed for writing in August of 2001, the same document previously states that the use of Item Response Theory enables comparisons to be made between the 2000 and 2001 results (British Columbia Ministry of Education, 2001).

This is not to suggest that changes to assessment policies or examinations are necessarily negative. Nor should the problems described above be viewed as isolated instances resulting from inexperienced Ministry personnel or poorly conceptualized assessment programs. In its review of assessment programs in Canada, the GAO (Government Accounting Office) applauded Canadian assessment practices, singling out Alberta and British Columbia as having exemplary programs (Cheliminsky & York,

1994). In particular, the GAO concluded that the involvement of teachers and the close fit between curriculum and the assessments were important factors in the successful implementation and use of the assessments. Certainly, the British Columbia Ministry of Education has modified assessment policies in response to teacher concerns. For example, shortly after reintroducing the provincial examinations in 1984, the value of the examinations was reduced from 50% to 40%. This small modification, viewed as a victory by teachers advocating fairness for students, may have helped to reduce some of the concerns teachers had regarding the examination program. Similarly, the inclusion of the second writing item and the use of the performance standards to assess the FSA writing have been viewed positively since they address the concerns and incorporate the work of practicing teachers.

Changes and modifications to assessment programs are inevitable. Nonetheless, the impact of such changes should not result in unfair and negative consequences for students, nor should such changes lead to unreasonable comparisons or policy decisions. The impact of the change in policy resulting from the removal of the scholarship examinations was foreseeable. A preliminary analysis of the results before the change in policy would have enabled Ministry officials to examine alternatives that would have reduced the false negative decisions associated with the current procedure. Similarly, the problems associated with the FSA could have largely been avoided if 1) the initial administration of the assessment had been delayed while reporting procedures were finalized; 2) more extensive field testing of the writing assessment and its associated scoring procedure had been completed; and 3) clear statements were provided by the

ministry of education regarding the types of analyses and comparisons that could be made based on the results.

Recommendations

Given the general lack of expertise regarding the use and interpretation of assessment information, it is important that the foundations for assessment be strongly conceptualized and the need for subsequent changes minimized. As the expansion of large scale assessment programs continues across Canada, those responsible for the implementation of these programs need to ensure that sufficient time is provided to develop assessment programs that meet their stated goals when they are implemented. Work is also required to support the proper use and interpretation of these results. In particular, the following recommendations, while not exhaustive, serve to provide guidelines for assessment practices that will minimize the need for future modifications and the associated impacts of change:

1) Given the expense of sound assessments, carefully consider the need for such new or expanded assessments.
2) Be honest with respect to the use of assessment programs and implement procedures to prevent the misuse of the results.
3) Develop a sound implementation plan, and provide time to implement the assessment program in a defensible manner.
4) Involve key stakeholder groups.
5) Identify potential problems before rather than after the fact.
6) Communicate results in a clear unambiguous manner.
7) Do not allow reduced costs to be a factor for changing assessment practices.
8) Carefully consider the range of implications associated with changes to existing assessment programs.
9) If changes are made, examine procedures to maximize comparability or develop policies to prevent inappropriate comparisons.

The importance and use of large-scale assessments is increasing, resulting in the development and expansion of assessment programs. Given the consequences and influences such assessments have on students, teacher practices, and policy decisions, it

is vital that the assessment data provide accurate and reliable information. Changes to assessment programs can result in alternative decisions, decisions that may unfairly harm students or lead to misinformed policy directions. If assessment is to be a cost-effective and useful mechanism for educational decisions, such problems must be prevented as opposed to repaired.

References

Baker, E. L., O'Neil, H. F., & Linn, R. L. (1994). Policy and validity prospects for performance based assessment. *Journal for the Education of the Gifted. Special Issue: Reformers Speak,* 17, 331-353.

Bennett, R. E. (1998). *Reinventing Assessment: Speculations on the future of large-scale educational testing. A policy information perspective.* Princeton, NJ: Policy Information Center, Educational Testing Service.

Bolger, N. & Kellaghan, T. (1990) Method of measurement and gender differences in scholastic achievement. *Journal of Educational Measurement,* 27, 165-174.

Bridgeman, B. (1989). *Comparative validity of multiple choice and free-response items on the advanced placement examination in biology* (College Board Report No. 89-2). New York: College Entrance Examination Board. (ERIC Document Reproduction Service No. ED 308 228)

British Columbia Ministry of Education (October 2000). *British Columbia Foundation Skills Assessment 2000: Provincial Results Report.* Retrieved November 10, 2001 from http://www.bced.gov.bc.ca/assessment/fsa/2000results/fsareportfinal.pdf

British Columbia Ministry of Education (October 2001). *British Columbia Foundation Skills Assessment 2001: Provincial Results Report.* Retrieved November 10,2001 from http://www.bced.gov.bc.ca/assessment/fsa/results/2001/prov10.pdf

Chambers, J. M., Cleveland, W. S., Kleiner, B. & Tukey, P. A. (1983). *Graphical methods for data analysis.* Belmont CA: Wadsworth International Group; Boston: Duxbury Press.

Cheliminsky, E., & York, R. L. (1994). *Educational testing: The Canadian experience with standards, examinations, and assessments.* General Accounting Office Report PEMD-93-11. Gaithersburg: MD. GAO.

Delandshere, G. (2001). Implicit theories, unexamined assumptions and the status quo of educational assessment. *Assessment in Education, 8,* 113-133.

DeMars, C. E. (1998). Gender differences in mathematics and science on a high school proficiency exam: The role of response format. *Applied Measurement in Education,* 11, 279-299.

Garner, M., & Englehard, G., Jr. (1999). Gender differences in performance on multiple-choice and constructed response mathematics items. *Applied Measurement in Education,* 12, 29-51.

Henderson, D. L. (1999). *Investigation of differential item functioning in exit examinations across item format and subject area.* Unpublished doctoral dissertation, University of Alberta, Edmonton, Canada.

Kolen, M. J., & Brennan, R. L. (1996). *Test equating: Methods and practices.* New York: Springer.

Lafleur, C. & Ireland, D. (1999). *Canadian and provincial approaches to learning assessments and educational performance indicators.* Technical report submitted to Commonwealth Caribbean Program, Americas Branch: The Canadian International Development Agency.

Livingston, S. A.; & Zieky, M. J (1983). *A Comparative study of standard-setting methods.* Report: ETS-RR-83-38. Princeton, NJ: Education Testing Service.

Schmitt, A. P., Mazzeo, J., & Bleistein, C. (1991, April). *Are gender differences between Advanced Placement multiple choice and constructed response sections a function of multiple choice DIF?* Paper presented at the annual meeting of the National Council of Measurement in Education, Chicago, IL.

Shealy, R., & Stout, W. F. (1993). A model-based standardization approach that separates true bias/DIF from group difference and detects test bias/DIF as well as item bias/DIF. *Psychometrika*, 58, 159-194.

Thomasson, G. L. (1997). *The goal of equity within and between computerized adaptive tests and paper and pencil forms.* Paper presented at the annual meeting of the National Council of Measurement in Education, Chicago, IL.

Wainer, H. & Thissen, D. (1993). Combining multiple-choice and constructed-response test scores: Toward a Marxist theory of test construction. *Applied Measurement in Education*, 6, 103-118.

Yen, W. M., & Ferrara, S. (1997). The Maryland school performance assessment program: Performance assessment with psychometric quality suitable for high stakes usage. *Educational and Psychological Measurement*, 57, 60-84.

Table 1: RMSE and Scholarship Examination Classification Errors Rates using the Current Procedure

Subject	1994/1995					1995/1996				
	Number of Students	RMSE	False Negative	False Positive	Error Rate (%)	Number of Students	RMSE	False Negative	False Positive	Error Rate (%)
Biology (Jan.)	1587	48.3	10.5	0.5	11.0	1867	40.8	7.1	1.4	8.5
Biology (June)	3227	42.6	7.0	2.0	8.9	3604	38.7	5.7	2.4	8.1
Chemistry (Jan.)	1534	58.4	15.6	0.2	15.8	1833	47.5	7.1	2.3	9.4
Chemistry (June)	3913	44.0	7.7	1.6	9.3	4170	45.7	8.3	2.0	10.3
Fr. Langue (June)	473	74.7	19.9	6.1	26.0	631	77.7	21.4	2.5	23.9
French (Jan.)	1263	52.3	12.4	1.1	13.5	1403	52.2	8.3	2.3	10.5
French (June)	--	--	--	--	--	2576	50.4	11.1	1.2	12.4
Geography (Jan.)	854	71.7	18.9	2.1	20.6	1052	74.0	19.4	3.3	22.7
Geography (June)	1717	64.4	15.0	3.4	18.5	1904	62.9	10.9	3.7	14.5
Geology (June)	306	57.0	9.8	3.6	13.4	308	64.4	19.5	1.3	20.8
History (Jan.)	998	77.1	20.3	2.0	22.3	1159	67.7	19.7	1.5	21.1
History (June)	2259	65.9	18.5	1.8	20.3	2017	62.8	12.4	5.0	17.4
Literature (Jan.)	446	80.4	24.2	0.9	25.1	252	66.8	16.2	2.4	18.6
Literature (June)	1512	68.5	19.9	1.7	21.6	725	59.3	10.6	5.6	16.2
Mathematics (Jan.)	2259	51.4	10.8	0.7	11.5	2765	55.3	13.9	1.4	15.3
Mathematics (June)	5283	51.4	8.9	0.9	9.8	5252	52.7	10.3	1.6	11.9
Mandarin (June)	543	63.2	5.0	12.5	17.5	775	66.3	4.0	8.9	12.9
Physics (Jan.)	772	64.34	16.5	1.0	17.5	945	55.84	12.8	0.8	13.7
Physics (June)	2601	48.40	6.8	3.5	10.4	2860	47.49	11.6	2.2	13.8
Total	5009	--	5.0	2.9	7.9	6069	--	5.1	3.4	8.4

Note: Differences in the overall error rate are due to rounding