4. STANDARD SETTING AND MAINTENANCE

This chapter gives an explanation of how the Basic Competency (BC) standards were set and maintained as well as how students' ability indices were estimated. It also summarizes the territory-wide results of the TSA 2014.

How the BC Standards were Set

Basic Competencies are the essential knowledge/skills required by students in relation to the learning targets and objectives set out in the curriculum for each key stage (P.3, P.6 and S.3). The Basic Competencies represent just part of the curriculum requirements. After the first year's administration of the TSA at each level (i.e. P.3 in 2004, P.6 in 2005 and S.3 in 2006) by the HKEAA, panels of judges were formed to set the BC standards for the three subjects: Chinese Language, English Language and Mathematics. The BC standards set remain unchanged across the years.

Two well-known methodologies, namely the Angoff method and the Bookmark method, were used for setting the standards. For the Angoff method, the judges were asked to imagine a 'minimally acceptable student' at the end of his/her respective key stage (P.3, P.6 or S.3) who has just grasped the Basic Competencies. Each judge was asked to write down in a well prepared form their envisaged probabilities for this student to answer each of the items correctly. The average of the totals of these probabilities of the entire panel, excluding the outliers, would be compiled. For the Bookmark method, each judge was required to insert a metaphorical 'bookmark' in the pile of a sample of scripts/ performances to separate those deemed as meeting the standard and those not meeting the standard. The results of this exercise, excluding those of the lenient and inconsistent judges, were pooled and a consensus judgment made about the final position of the 'bookmark'. The results of these two methods were considered alongside relevant international standards in determining the final cut scores. This ensures that the standards set in Hong Kong are competitive with those of other regions.

How the BC Standards are Maintained

To maintain the standards set, a research test (or anchor test) is used to link and equate students' performance shortly before the conduct of each year's TSA. This research test was taken by a specified number of students on a stratified sampling basis in the first year (Year 1 in Table 4.1) when it was approaching to the assessment dates of the TSA. In the

subsequent year (Year 2 in Table 4.1), the same test was taken by about the same number of students as in Year 1 close to the implementation of the TSA. Table 4.1 shows how students' responses data are linked into a big matrix.

Table 4.1 Linking Methods in Standard Maintenance

Item / Student	TSA Year 1	Research Test	TSA Year 2
Students in Year 1	Students' responses in Year 1	Students' responses in Year 1	
Students in Year 2		Students' responses in Year 2	Students' responses in Year 2

In Year 1, the difficulty indices of the research test items would be estimated together with that of the TSA items. Similarly in Year 2, the difficulty indices of the research test items would also be estimated together with that of the TSA items. By assuming the difficulty indices of the research test items being comparable, the difficulty indices of the TSA items in Year 2 could be calibrated with Year 1. In other words, with the common research test, the difficulty indices of the TSA items in Year 1 and Year 2 could be calibrated on the same scale. Hence, the performance of the students in Year 2 is comparable to that of the students in Year 1. The benchmark set in the first year's TSA (i.e. P.3 in 2004, P.6 in 2005 and S.3 in 2006) could then be used to determine which students in the subsequent years can achieve the BC standard. In doing so, the benchmark of the BC standard set in the first year remains unchanged across the years.

Estimate Students' Ability Indices

For each of the three subjects (namely Chinese Language, English Language and Mathematics) in the TSA, one single paper which covers the full BC scope would be too lengthy for a student. Therefore, several sub-papers would be set for each subject where a student is only required to attempt one of the sub-papers. There would be a number of overlapping items covered among the sub-papers for equating purposes. Table 4.2 is an illustrative example of the paper design for a subject for all students on three sub-papers.

Table 4.2 Overlapping Items in Paper Design

Item / Sub-paper	1	2	3	4	5	6
Sub-paper 1	■	■	■			
Sub-paper 2			■	■	■	
Sub-paper 3		■			■	■

After administrating the assessment, the responses from all students of the three sub-papers are merged into a single data matrix from which the item difficulty indices as well as students' ability indices are estimated using psychometric methods. Since each sub-paper includes overlapping items for equating purposes, a student's ability index can be estimated regardless of the difficulty of the sub-papers. In other words, the measure of a student's ability index is independent of which sub-paper he/she attempts.

Results of TSA 2014

The aforementioned procedures for standard maintenance are applied to calculate percentages of students achieving basic competency in TSA 2014. The final result in the territory-wide percentages of students achieving basic competency in 2014 is summarized in Table 4.3.

Table 4.3 Territory-wide Percentages of Students Achieving Basic Competency

Subject and Level		Percent Achieving Basic Competency										
		2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Chinese Language (Listening, Reading and Writing)	P.3	82.7	84.7	85.2	84.9	85.4	#	85.9	86.4	86.1	86.6	86.3
	P.6	--	75.8	76.5	76.7	76.4	#	77.0	77.2	^	78.1	^
	S.3*	--	--	75.6	76.2	76.5	76.5	76.8	76.7	76.9	77.1	77.0
English Language (Listening, Reading and Writing)	P.3	75.9	78.8	79.4	79.5	79.3	#	79.2	79.8	79.7	80.4	80.3
	P.6	--	70.5	71.3	71.3	71.5	#	71.6	71.7	^	72.4	^
	S.3	--	--	68.6	69.2	68.9	68.8	69.2	69.2	69.1	69.5	69.3
Mathematics	P.3	84.9	86.8	86.9	86.9	86.9	#	87.0	87.0	87.3	87.5	87.4
	P.6	--	83.0	83.8	83.8	84.1	#	84.2	84.1	^	84.2	^
	S.3	--	--	78.4	79.9	79.8	80.0	80.1	80.1	79.8	79.7	79.9

Note: * Chinese Audio-visual component included in the calculation of the cut score at the S.3 level since 2007.

Due to Human Swine Influenza causing the suspension of primary schools, the TSA was cancelled and no data has been provided.

^ The 2012 and 2014 P.6 TSA were suspended. As participation in the 2012 and 2014 P.6 TSA was on a voluntary basis, not all P.6 students were involved and hence no territory-wide data is provided in this report.

On the whole, the proportion of students achieving basic competency at P.3 and S.3 was highest in Mathematics followed by Chinese Language and English Language. Table 4.3 shows the proportion of students achieving basic competency decreases over the Key Stages. Examining the performance of P.3 and S.3 students, it is possible to discern overall trends, which are shown graphically in Figures 4.1 and 4.2.

Figure 4.1 P.3 Territory-wide Percentages of Students Achieving Basic Competency

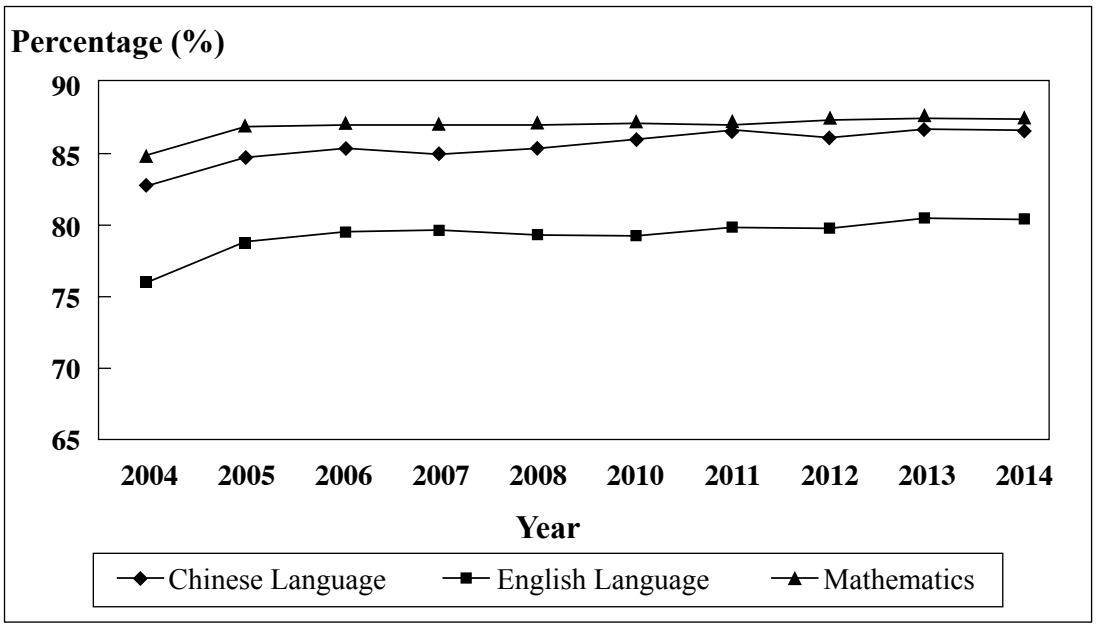

Figure 4.2 S.3 Territory-wide Percentages of Students Achieving Basic Competency

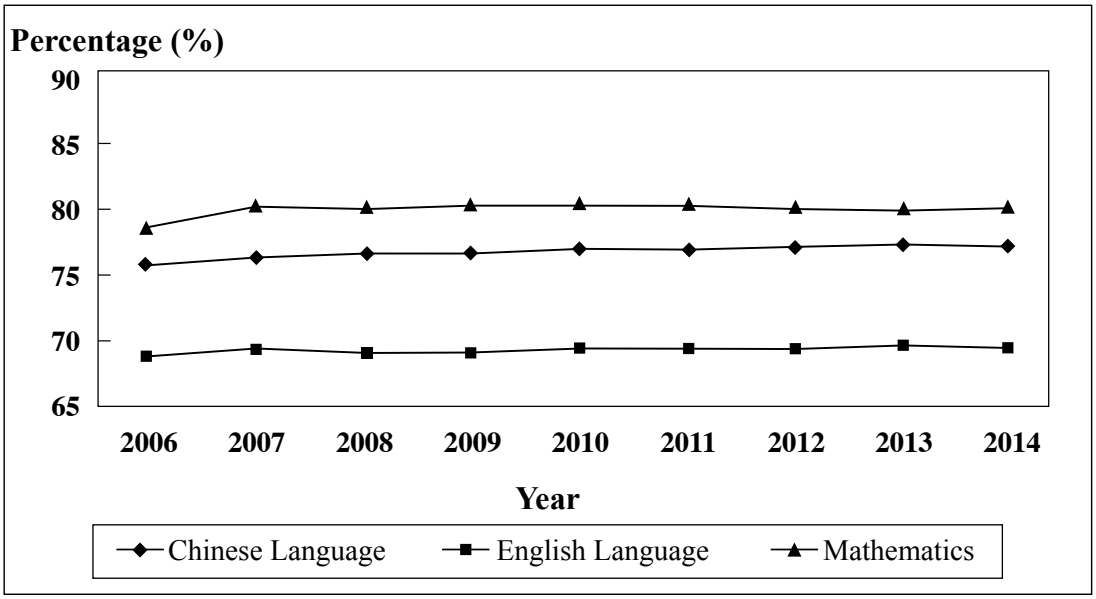

Table 4.4 summarizes some key statistics for those TSA 2014 students who also took the TSA three years ago.

Table 4.4 Number and Percentages of Cohort Students Achieving or Not Achieving Basic Competency in 2011 P.6 and 2014 S.3

Subject	Chinese Language	English Language	Mathematics
Achieved both P.6 BC in 2011 and S.3 BC in 2014	34,207 (71.4%)	32,502 (67.4%)	37,588 (77.9%)
Achieved P.6 BC in 2011 but not S.3 BC in 2014	4,225 (8.8%)	3,219 (6.7%)	4,217 (8.7%)
Achieved S.3 BC in 2014 but not P.6 in 2011	3,730 (7.8%)	2,393 (5.0%)	1,935 (4.0%)
Number of students sitting both P.6 TSA in 2011 and S.3 TSA in 2014	47,916	48,231	48,229

To generate the above table, it was necessary to link the data for 2011 and 2014. After matching the student records, approximately 48,000 students sat the P.6 TSA in 2011 and the S.3 TSA in 2014. As anticipated, most students who achieved basic competency in 2011 also achieved basic competency in 2014. These results indicate that having a solid learning foundation in junior levels is beneficial to learning in the next key stage. In addition, teachers' early acquisition of solid assessment data is most useful in enhancing students' learning.