4. STANDARD SETTING AND MAINTENANCE

This chapter sets out how the BC standards were set and maintained in the TSA as well as how students' ability indices were estimated. It also summarises the results of the 2018 TSA.

How the Standards were Set

BCs are the essential knowledge and skills to be acquired by students (only including part of knowledge and ability) in the three subjects of Chinese Language, English Language and Mathematics by the end of each key stage of learning (P.3, P.6 and S.3) as set out in the curriculum. After the first year's administration of the TSA for each level (i.e. P.3 in 2004, P.6 in 2005 and S.3 in 2006) by the HKEAA, expert panels were formed to set the BC standards for the three subjects of 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 experts were asked to imagine a student who has grasped the BCs at the end of his/her respective key stage (P.3, P.6 or S.3). Each expert was asked to write down 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 expert was required to insert a metaphorical "bookmark" in the pile of sample scripts to separate the performances of those deemed as meeting the standard and those not meeting the standard. The results of this exercise, excluding those of the lenient and inconsistent experts, were pooled and a consensus judgement 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 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 approaching the assessment dates of 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	Sample Students' Responses	
Students in Year 2		Sample Students' Responses	Students' Responses

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's. 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), 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 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. Under the new arrangements for the P.3 2018 TSA, the HKEAA continued to adopt the aforementioned methodology for standard setting and maintenance, as well as estimating students' ability indices.

Results of Territory-wide System Assessment in 2018

The aforementioned procedures for standard maintenance were applied and the final results in the percentages of P.3 and S.3 students achieving BCs in 2018 are summarised in Table 4.3.

Table 4.3 Territory-wide Percentages of P.3 and S.3 Students Achieving BCs

Subject and Level		Percentages of Students Achieving BCs														
		2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
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	86.4	85.8$^{\Delta}$	86.3$^{\triangledown}$	86.7$^{\square}$
	P.6	--	75.8	76.5	76.7	76.4	#	77.0	77.2	^	78.1	^	77.7	^	78.3	^
	S.3*	--	--	75.6	76.2	76.5	76.5	76.8	76.7	76.9	77.1	77.0	77.2	77.4	77.1	76.9
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	80.4	81.1$^{\Delta}$	81.1$^{\triangledown}$	80.8$^{\square}$
	P.6	--	70.5	71.3	71.3	71.5	#	71.6	71.7	^	72.4	^	72.0	^	72.3	^
	S.3	--	--	68.6	69.2	68.9	68.8	69.2	69.2	69.1	69.5	69.3	69.4	69.6	69.7	69.8
Mathematics	P.3	84.9	86.8	86.9	86.9	86.9	#	87.0	87.0	87.3	87.5	87.4	87.6	89.9$^{\Delta}$	88.2$^{\triangledown}$	88.0$^{\square}$
	P.6	--	83.0	83.8	83.8	84.1	#	84.2	84.1	^	84.2	^	84.0	^	84.0	^
	S.3	--	--	78.4	79.9	79.8	80.0	80.1	80.1	79.8	79.7	79.9	79.9	80.0	79.9	80.0

Note: * Chinese Audio-visual component has been 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 was provided.

^ The P.6 TSA was suspended in 2012 and 2014. Since 2015, the P.6 TSA has been implemented in odd-numbered years. School participation has been on a voluntary basis in even-numbered years. Since participation in this assessment was on a voluntary basis and not all P.6 students were involved, no territory-wide data is provided in this report.

$^{\Delta}$ The 2016 P.3 level assessment was conducted as part of the 2016 Tryout Study. The BC attainment rates of the Chinese Language, English Language and Mathematics subjects were calculated using the data from some 50 participating schools.

$^{\triangledown}$ The 2017 P.3 level assessment was conducted as part of the 2017 Research Study, which was extended to all primary schools in the territory.

$^{\square}$ Starting from 2018, the P.3 TSA is conducted on a sampling basis. The BC attainment rates are inferred from the sample of all students participating in the assessment.

The overall attainment rates of P.3 students in the subjects of Chinese Language, English Language and Mathematics were 86.7%, 80.8% and 88.0% respectively. For S.3, the attainment rates in the Chinese Language, English Language and Mathematics subjects were 76.9%, 69.8% and 80.0% respectively. On the whole, the proportion of students achieving BCs at P.3 and S.3 was highest in Mathematics, followed by Chinese Language and English Language. 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 BCs

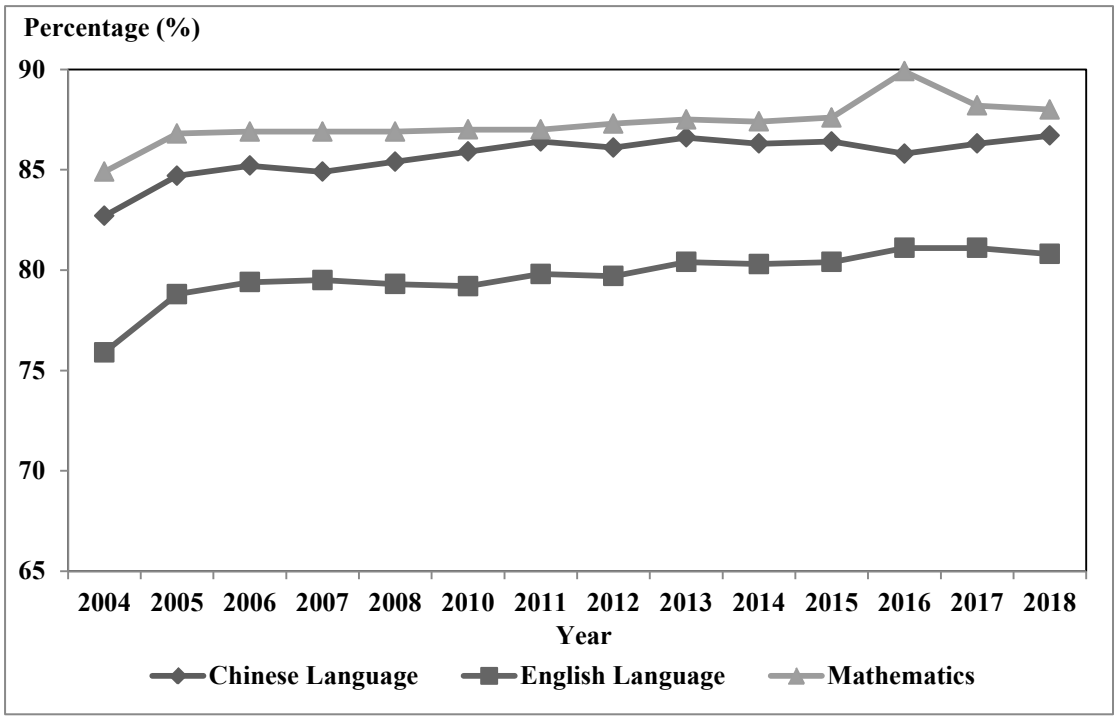

Figure 4.2 S.3 Territory-wide Percentages of Students Achieving BCs

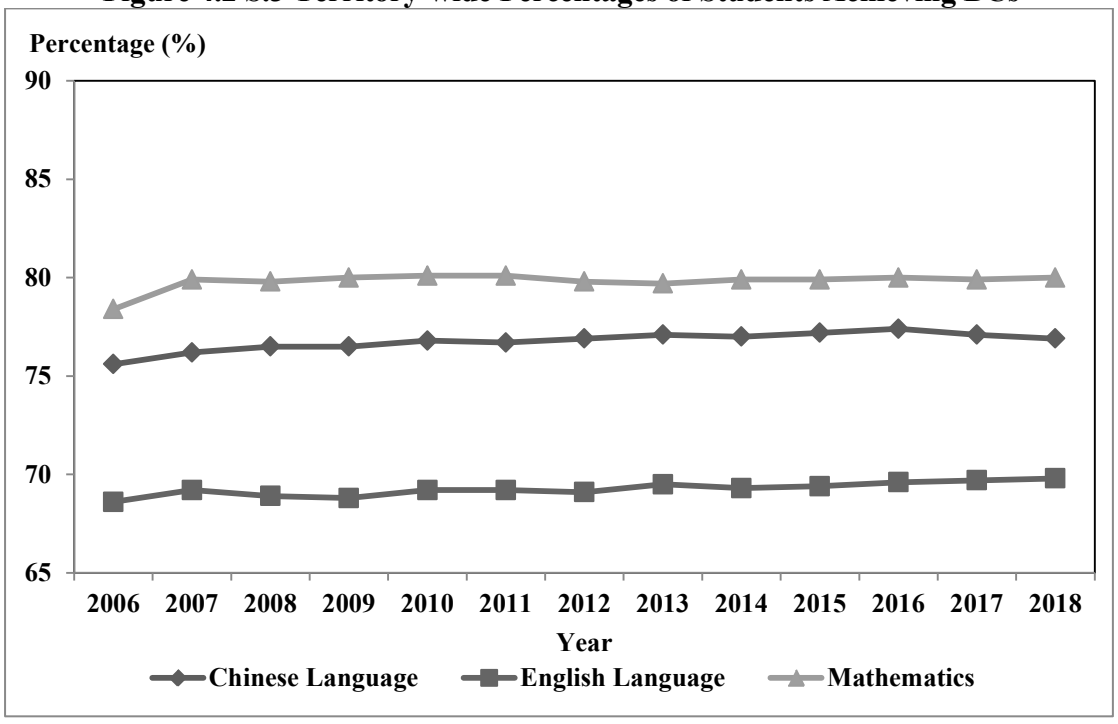

Table 4.4 summarises some key statistics for those 2018 TSA students who also took TSA three years ago.

Table 4.4 Number and Percentages of Cohort Students Achieving or Not Achieving BCs in 2015 P.6 and 2018 S.3

Subject	Chinese Language	English Language	Mathematics
Achieved both P.6 BCs in 2015 and S.3 BCs in 2018	28,647 (72.4%)	26,721 (67.8%)	30,855 (78.0%)
Achieved P.6 BCs in 2015 but not S.3 BCs in 2018	3,282 (8.3%)	2,522 (6.4%)	3,460 (8.8%)
Achieved S.3 BCs in 2018 but not P.6 in 2015	2,965 (7.5%)	2,485 (6.3%)	1,775 (4.5%)
Number of students sitting both P.6 TSA in 2015 and S.3 TSA in 2018	39,542	39,419	39,539

To generate the above table, it was necessary to link the data for 2015 and 2018. After matching the student records, over 40,000 students sat the P.6 TSA in 2015 and the S.3 TSA in 2018. Most students who achieved BCs in 2015 also achieved BCs in 2018. These results indicate that having a solid learning foundation at the junior level is beneficial to learning in the next key stage. It is important for teachers to obtain assessment data in enhancing students' learning.