This chapter sets out how the BC standards have been set and maintained in the TSA as well as how students' ability indices have been estimated. It also summarises the results of the TSA 2025.

Setting the Standards

BCs are the essential knowledge and skills (only including part of knowledge and skills in respective curriculum) to be acquired by students 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. In 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.

The Angoff method and the Bookmark method were used by the HKEAA for setting the standards. For the Angoff method, the experts exercised their own professional judgement to simulate the probability of answering each item correctly by "a minimally competent student meeting the BC standard". The results of each expert were pooled and revised before a consensus was reached on the final score of this student. For the Bookmark method, each expert inserted a metaphorical "bookmark" in the pile of sample scripts to segregate the performances of those deemed as meeting the standard and those not meeting the standard. The results of each expert were pooled and a judgement was made about the final position of the "bookmark" to indicate the relative BC standards. The results of these two methods were then considered alongside relevant international standards in determining the final cut scores to ensure the standards set in Hong Kong are comparable with those of other countries.

Maintaining the Standards

To maintain the pre-set BC standard, a Research Test (RT) is used by the HKEAA to link and equate students' performance between years. Students' performance in the current year is compared with that of the previous year, thus ensuring the stability and consistency of the standards. The method is detailed as follows: The RT was taken before the conduct of the TSA in the first year by a specified number of students selected on a stratified sampling basis. This group of students must also participate in the TSA that year. In the subsequent year, roughly the same number of sampled students took the same RT and the TSA of that year. Table 4.1 shows how students' response data in the RT and TSA in these two years are linked into a huge matrix.

Table 4.1 Design Pattern 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 those of the TSA items. Similarly in Year 2, the difficulty indices of the same research test items would also be estimated together with those of the TSA items. The difficulty indices of the TSA items in different years could be calibrated on the same scale. In other words, the performance of the students in different years could be comparable on the same scale. Hence, 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 whether students in the subsequent years can achieve the BC standard. In view of the above procedures, the BC standard set in the first year remains unchanged across the years.

Estimating Students' Ability Indices

For any one of the subjects of Chinese Language, English Language and Mathematics, a single paper which covers all BCs would be too lengthy with regard to paper design. Therefore, several sub-papers are set to cover all BCs, with each student required to attempt only 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 of TSA for a subject.

Table 4.2 Paper Design (With Overlapping Items)

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

After the assessment, the responses from all students on different 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 different sub-papers attempted by students. Therefore, the estimation of a student's ability index is independent of the difficulty of the sub-paper he/she attempts.

Results of Territory-wide System Assessment in 2025

The aforementioned procedures for standard maintenance were applied and the final results in the percentages of P.3, P.6 and S.3 students achieving BCs in 2025 are summarised in Table 4.3.

Table 4.3 Territory-wide Percentages of P.3, P.6 and S.3 Students Achieving BCs**

Subject and Level		Chinese Language (Listening, Reading & Writing)			English Language (Listening, Reading & Writing)			Mathematics		
		P.3	P.6	S.3*	P.3	P.6	S.3	P.3	P.6	S.3
Percentages of Students Achieving BCs	2004	82.7	--	--	75.9	--	--	84.9	--	--
	2005	84.7	75.8	--	78.8	70.5	--	86.8	83.0	--
	2006	85.2	76.5	75.6	79.4	71.3	68.6	86.9	83.8	78.4
	2007	84.9	76.7	76.2	79.5	71.3	69.2	86.9	83.8	79.9
	2008	85.4	76.4	76.5	79.3	71.5	68.9	86.9	84.1	79.8
	2009	#	#	76.5	#	#	68.8	#	#	80.0
	2010	85.9	77.0	76.8	79.2	71.6	69.2	87.0	84.2	80.1
	2011	86.4	77.2	76.7	79.8	71.7	69.2	87.0	84.1	80.1
	2012	86.1	^	76.9	79.7	^	69.1	87.3	^	79.8
	2013	86.6	78.1	77.1	80.4	72.4	69.5	87.5	84.2	79.7
	2014	86.3	^	77.0	80.3	^	69.3	87.4	^	79.9
	2015	86.4	77.7	77.2	80.4	72.0	69.4	87.6	84.0	79.9
	2016	85.8$^{\Delta}$	^	77.4	81.1$^{\Delta}$	^	69.6	89.9$^{\Delta}$	^	80.0
	2017	86.3$^{\nabla}$	78.3	77.1	81.1$^{\nabla}$	72.3	69.7	88.2$^{\nabla}$	84.0	79.9
	2018	86.7$^{\square}$	^	76.9	80.8$^{\square}$	^	69.8	88.0$^{\square}$	^	80.0
	2019	85.8$^{\square}$	77.9	76.4	79.8$^{\square}$	72.8	69.5	87.7$^{\square}$	84.2	79.6
	2023	82.4$^{\square}$	71.1	74.7	79.5$^{\square}$	64.3	67.8	86.5$^{\square}$	78.3	76.6
	2024	80.9$^{\square}$	^	77.0	78.7$^{\square}$	^	67.0	85.3$^{\square}$	^	79.0
	2025	81.4$^{\square}$	70.5	78.0	83.2$^{\square}$	70.2	68.8	85.4$^{\square}$	79.0	78.2

Notes:* 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 2009 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.

$^{\nabla}$ 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.

** Due to the volatility of the COVID-19 epidemic, the TSA 2020, 2021 and 2022 were suspended and no data was provided.

The overall attainment rates of P.3 students in the subjects of Chinese Language, English Language and Mathematics were 81.4%, 83.2% and 85.4% respectively. For P.6, the attainment rates in the subjects of Chinese Language, English Language and Mathematics were 70.5%,

70.2% and 79.0% respectively. For S.3, the attainment rates in the subjects of Chinese Language, English Language and Mathematics were 78.0%, 68.8% and 78.2% respectively. The overall performance trends of P.3, P.6 and S.3 students are shown in Figures 4.1, 4.2 and 4.3.

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

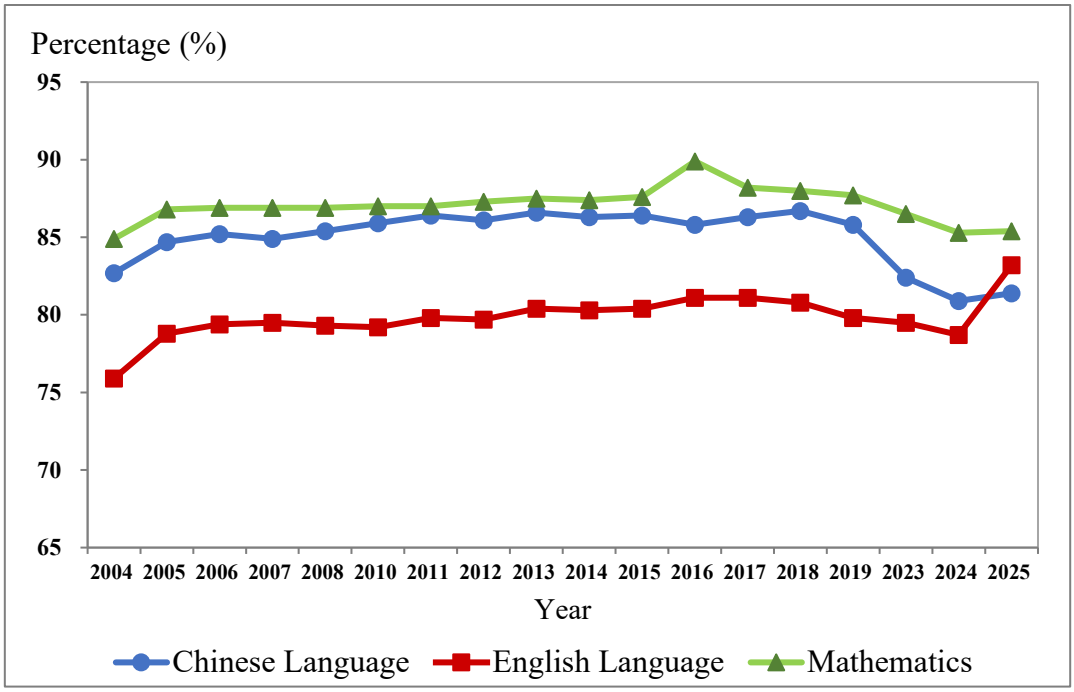


Figure 4.2 P.6 Territory-wide Percentages of Students Achieving BCs

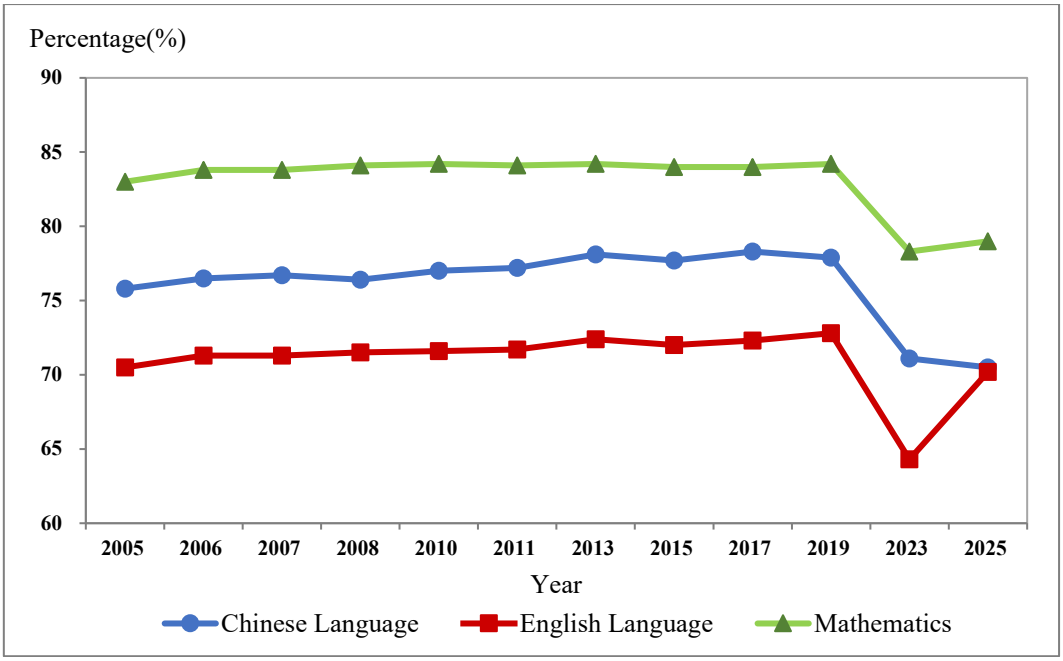
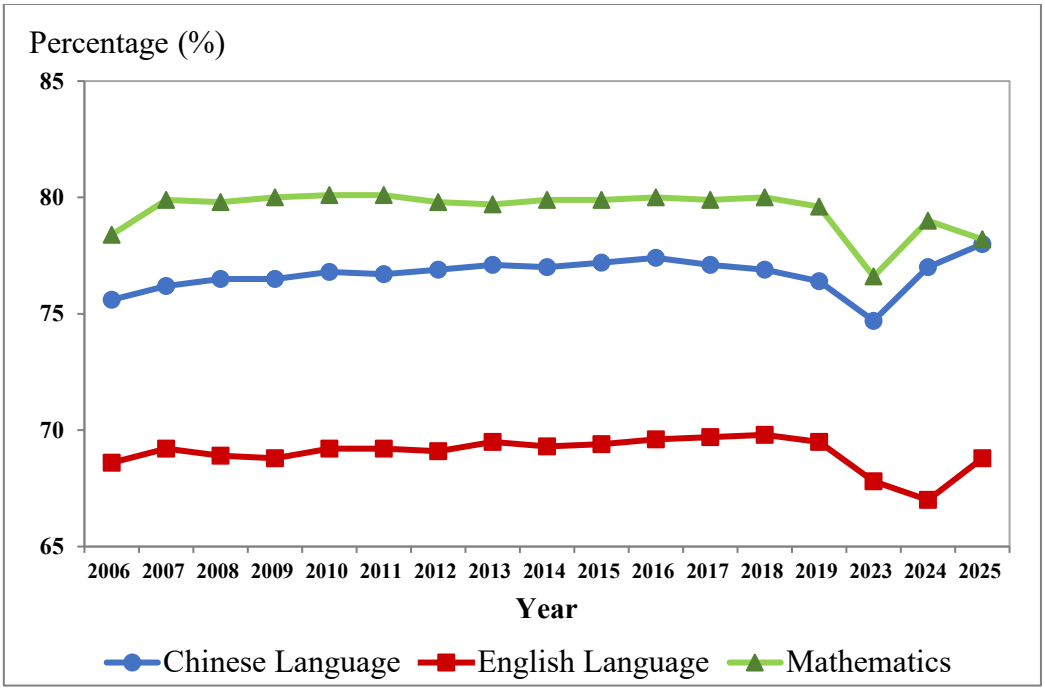


Figure 4.3 S.3 Territory-wide Percentages of Students Achieving BCs



In general, the prerequisite for the HKEAA to provide the number and percentage of cohort students achieving or not achieving BC in the P.3 and P.6 TSA, as well as the P.6 and S.3 TSA is that the P.3 and P.6 students, as well as the P.6 and S.3 students participating in the respective TSA should be of the same cohort. In accordance with the above principle, starting from 2018, the P.3 TSA has been conducted on a sampling basis, and hence from 2019 onwards, the HKEAA cannot match the records of the same cohort of P.3 and P.6 students. The TSA 2022 was suspended due to the COVID-19 epidemic. Consequently, this year's report cannot provide the record-matching data for the corresponding cohort of P.6 and S.3 students.