

## **Adapting the Cognitive Abilities Test (CAT4) to support decision making, teaching and learning in Chinese schools**

**Keywords: Test development, Digital, ability assessment**

Bernadetta Brzyska, Head of Research, GL Education,

[Bernadetta.brzyska@gl-assessment.co.uk](mailto:Bernadetta.brzyska@gl-assessment.co.uk)

### Abstract

GL Assessment's Cognitive Abilities Tests (4<sup>th</sup> edition) (CAT4) is taken by thousands of children every year to provide an understanding of student potential. It identifies their strengths, weaknesses and thinking preferences, providing accurate and reliable information for teaching and learning. Schools are provided with automatically generated reports that offer detailed 'learner profiles', helping teachers to shape their teaching for each student, making informed decisions about how to support their students in the classroom.

The theme of this year's IAEA conference focuses on Assessment and Decision making across individual and institutional solutions. It is therefore timely to present our study (GL Assessment, 2018) showing the findings of the nationwide trial of a Chinese version of CAT4 in China including a standardisation of 15,000 students, and how we aim to encourage the use of CAT4 for improved educational outcomes. Individualised or personalised learning is still a relatively new concept for teachers in many Chinese schools and China has started moving away from passive and rote learning style to a more "active, problem-solving learning style to improve students' overall abilities to process information, acquire knowledge, solve problems and learn cooperatively" (OECD 2016).

To make the test culturally and linguistically suitable for Chinese students, the test needed to be translated and adapted, replacing and amending existing items. For those interested in adapting and translating tests into other languages and contexts, this paper will also detail some of the issues and barriers encountered in developing a digital assessment in China.

A study of this size focusing on reasoning abilities will be of interest to test researchers and educationalists who would like to understand more about the use of ability tests in school age children in China – a key assessment for decision making at individual and institutional level.

## **Introduction**

GL Education's <sup>1</sup> Cognitive Abilities Test: 4<sup>th</sup> Edition (CAT4) is a standardised assessment testing cognitive abilities in children aged between 5 and 18. CAT4 is used with children worldwide – in both the UK and international school sectors where English is the main or one of the main modes of instruction. CAT4 data is often considered alongside other educational data such as attainment and attitudinal information.

The theme of this year's IAEA conference focuses on Assessment and Decision making across individual and institutional solutions. It is therefore felt to be an ideal opportunity to share an overview of the CAT4, alongside highlights from GL Education's recent report (2018) detailing the adaptation of the test into Chinese, and how we aim to support schools who will be able to use CAT4 to inform decisions at both individual and group level.

### **How is CAT4 used?**

CAT4 is widely used to provide an understanding of student potential, informing teachers on their students' ability, incorporating verbal, spatial, non-verbal and quantitative reasoning. As such, CAT4 reports provide teachers with recommendations for teaching and learning, based on each individual student's results.

For teachers, having a clear understanding of a child's abilities can support them in their teaching practice in many ways, allowing them to identify children's strengths, weaknesses and thinking preferences. At the institutional or classroom level teachers in the UK use CAT4 to plan teaching and curriculum, monitor attainment and set aims and aspirations. CAT4 can be used to look at a whole group of students, comparing their standardised scores against the national population. It is also possible to explore subgroups of students, identifying those who are most able, highlighting uneven learning profiles, or uncovering possible support needs that might be more difficult to discern quickly in the classroom. The reports also help schools to identify children who may be gifted and talented, those who may be under-achieving, as well as those who may have unrecognised special educational needs.

For many children, CAT4 reveals hidden abilities and changes both the student's and the teacher's perceptions and approach to each child's learning.

More information about CAT4 reports, including sample reports is available on the GL Education website<sup>2</sup>, alongside school case studies showing how teachers in the UK and internationally use CAT4.<sup>3</sup>

### **Education in China**

China has the largest education system in the world, with almost 260 million students and over 15 million teachers in 514,000 schools (National Bureau of Statistics of China, 2014).

Individualised or personalised learning is still a relatively new concept for teachers in many Chinese schools and China has started moving away from passive and rote learning style to a more "active, problem-solving learning style to improve students' overall abilities to process information, acquire knowledge, solve problems and learn cooperatively" (OECD 2016). Indeed, this was one of the six main objectives of the Basic Education Curriculum Reform Outline (Cui et al, 2018) in 2001. Another objective (#5) also outlines the need for curriculum evaluation to move from narrowly summative to more formative purposes, away

---

<sup>1</sup> [www.gl-education.com](http://www.gl-education.com) - a provider of formative assessment and benchmarking in the UK

<sup>2</sup> Further information on CAT4 and its reporting system can be found at [www.gl-education.com/support/cat4-product-support/cat4-reports](http://www.gl-education.com/support/cat4-product-support/cat4-reports)

<sup>3</sup> <https://www.gl-assessment.co.uk/news-hub/case-studies/?assessments=CAT4&page=1>

from screening and selection, and towards a promotion of student development through good teaching practice and instructional improvement (Cui, Lei & Zhou, 2018; OECD, 2016). There is a need to look at a diverse range of criteria for student outcomes, changing the examination-oriented study mode to reduce the burden on students. Chinese schools now need assessments that help to students and their teachers to discover and develop their potentials, as well as understand themselves and build self-confidence (OECD, 2016).

These changes in China gave an opportunity for GL Education and partners to explore the possibility of developing a Mandarin Chinese version of CAT4, providing a new formative tool for teachers and students to use in Chinese schools. In September 2016, GL Education started a partnership with China Future School Lab (National Institute of Education Sciences), engaging in a study to adapt CAT4 and produce a norm referenced dataset to reflect China’s education system and its students.

### **CAT4 technical information**

To understand the process of adapting CAT4 into Chinese, first we must cover the technical aspects of CAT4. CAT4 originated from items in the CogAT (a US test battery), which led to the creation of UK standardised CAT3 (Smith, Fernandes & Strand, 2003). This later became CAT4 with the addition of the Spatial battery, the development of which was led by developer and psychologist, Dr Pauline Smith. The history of CAT4 is beyond the scope of this paper, but more information about its development and UK standardisation can be found in the technical manual (GL Assessment, 2012).<sup>4</sup>

Published in 2012, the assessment was developed in the medium of English and each level of the assessment is designed for a specific age group<sup>5</sup> (Figure 1), with some common items linking levels.

Level	Age range for general population norms
A	7:06–9:11
B	8:06–10:11
C	9:06–11:11
D	10:06–12:11
E	11:06–13:11
F	12:06–15:11
G	14:06–17:00+

Figure 1: Levels of digital CAT4 UK

CAT4 is available in both paper and digital formats and presents items in a fixed format, with easier items appearing first in a subtest. The test is timed and is generally administered within three time periods (around 40 mins each). It comprises of four separate batteries concentrating on verbal, spatial, non-verbal and quantitative reasoning. Each of the batteries can be broken down further into sub-sections (Figure 2).

<sup>4</sup> [https://www.gl-assessment.co.uk/media/1343/cat4\\_extended\\_technical\\_information.pdf](https://www.gl-assessment.co.uk/media/1343/cat4_extended_technical_information.pdf)

<sup>5</sup> Ages 5-7 are covered by paper only versions of the test

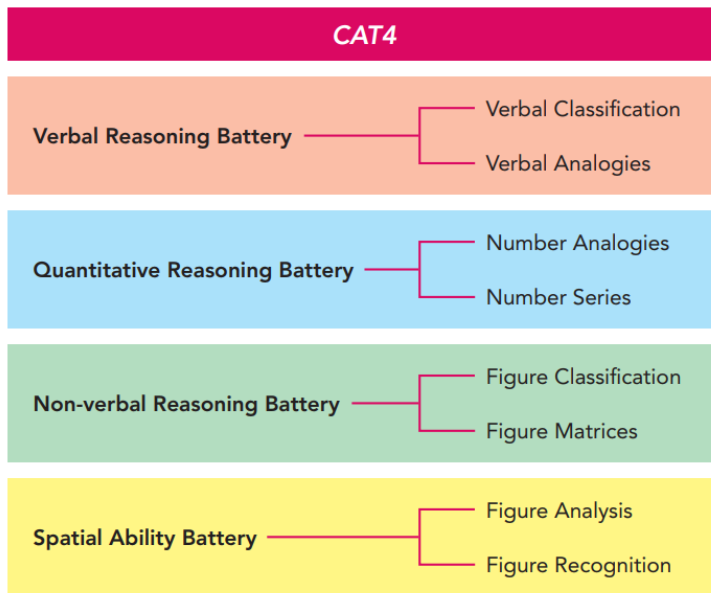


Figure 2: CAT4 batteries

Three of the four test batteries in CAT4 are not dependent on language (Quantitative, Non-Verbal and Spatial) – these batteries are successfully used by children who do not have English as a first language. Verbal Reasoning can relay depressed scores for these children due to their lower levels of English proficiency. International schools who teach in English often re-test with CAT4 over the course of a student’s academic career, to show improvements in verbal scores as the child learns and becomes more proficient in English. However, in other countries with students whose mode of learning is in other languages, this assessment is not deemed appropriate and so over the years there has been an increasing request for a version of CAT4 in other languages. A Welsh (2013)<sup>6</sup> and Spanish (2016)<sup>7</sup> version have been adapted providing GL Education’s first experience in researching CAT4 in different environments and languages. These two versions are outside the remit of this paper.

### Adapting for China

GL Education worked closely with two consultant developers, Dr Yan Tong and Ms Guo Jia Qi, who were first language mandarin speakers, fluent in English and understood the model of ability testing. Importantly, they were also close to the Chinese education system, and understood the differences in cultural and linguistic nuances of Chinese Mandarin across different regions of China. Over the course of the study, test content with translated instructions, examples and practice items was adapted and created to be made accessible to students in China.

With the help of these consultants, the Verbal battery in particular was carefully translated, adapted and reviewed. It is important to note that although CAT4 is available to a number of age groups, the remit of the study was to focus on adapting items from two levels of CAT4, covering levels E and F. This age group of children was chosen because, in most regions of China, children start secondary school at age 12 so this an important transition time for them from primary into secondary education.

### Item review and language differences

The consultants first translated, then reviewed all verbal items. Each consultant worked on one level each. Where there were common items across the two levels, this gave an

<sup>6</sup> Carried out in conjunction with Bangor University

<sup>7</sup> A small scale research project with a teacher in Argentina

opportunity for a comparison of the translation and interpretation. They were asked to amend and replace items where appropriate and communicated daily, working through issues together.

Overall approximately a third of the verbal items were amended or replaced after translation and this applied across both Verbal analogies and Verbal Classification sections. The review process highlighted some very interesting language differences and learning points. Some examples are below.

Some items needed a change due to differences in familiarity that were contextually specific to English. For example, one item had to change where the stem showed types of bread in the UK. In the Chinese version the whole item was changed to show types of tea as this was more equivalent in terms of familiarity. Another item, about flavours of ice cream, was felt to be equivalent, but the ice cream flavours themselves that made up the stem were changed to ones that were more familiar to Chinese children; for example, vanilla was changed to red bean. Sometimes only one of the words in the stem had to change, for example, some concepts of medical signs/symptoms/conditions were not thought to be known to Chinese children in the same way as UK children ('measles' which is more commonly known by English children was changed to 'Hepatitis'). This was also the same for the concept of 'attic' which was felt not to be as familiar to Chinese children many of whom live in apartments.

Some items did not work linguistically in Chinese, such as those which looked at groups or genders of animals or shades of colours. Sometimes a single distractor needed to be changed as the translation meant that the item no longer referenced the same linguistic rationale. For example, the item 'step stool' changed to 'step ladder' as the translation for step stool is "little stool" which would not hold the rationale of 'upward movement'.

In some cases, the English words which have several synonyms, did not translate into Chinese. For example, 'maggot' and 'grub' in Chinese are the same word once translated. Another item was replaced entirely because there were not enough stem words available for the word 'burn' in Chinese, where the English language has several synonyms.

Another item referred to the correct answer 'sibling' – where in Chinese the Chinese character contains the characters for 'son' and 'daughter'. 'Son' was one of the other distractors in the item and therefore the distractor had to be amended.

We found that nuances in the English language which are a play on words, don't work in Chinese. For example, in one question the stem related to parts of the eye including 'pupil'. One of the distractors in English was 'student', but this play on words between pupil (eye) and student did not work in Chinese. Similarly, the connection between the word 'lash' (which can mean an eyelash or to hit something) and the distractor 'whip' did not work in Chinese.

Once an external review was also carried out by China Future School Lab (National Institute for Education Sciences), including some Beijing based reviewers, amends were agreed, and the digital test was programmed and launched on a digital testing platform.

### **Standardisation**

The tests were taken in the summer term of 2017. With the help of our Chinese partners, overall, 15,000 students took part in the study<sup>8</sup> from 180 schools across 19 provinces in China. Further information on the sample, including how it was split by province, gender, urban/rural and ethnicity is available in the full research report (GL Assessment, 2018).

### **Reports and feedback from schools**

---

<sup>8</sup> CAT4 was originally standardised with over 25,000 children in the UK.

The UK version of CAT4 comes with several detailed reports for teachers, students, senior school leaders and even parents. As a first step, it was felt important that the priority for report development should focus on what could be used by teachers and students. Three reports were therefore adapted and developed for China: the group report for teachers, the individual student report for teachers and the individual report for students.

Chinese versions of the reports were developed keeping in mind a context that was meaningful for Chinese teachers and their students. Ultimately the aim was to encourage those teachers to use the assessment that can offer the same support, information and insights that flow from CAT4 in the UK. These reports cover a range of measures and important information including:

- An assessment overview with example questions and explanations behind each battery
- SAS scores and group ranking for a group of students
- Group scores by battery
- Student profiles – grouping those students with similar or contrasting profiles to be taught together. The section explains the general characteristics of each profile type, comparing group results to the national sample.
- An analysis of an individual pupil's results, along with a focus on how they can be helped to achieve their potential and actionable implications for teaching and learning provided for each pupil. This offers brief insights into how different levels of ability combined with learning preferences may affect a pupil's learning.
- Reports for students show them an explanation of their CAT4 results and where their strengths and weaknesses lie. It is important for all pupils to understand that the information gained from CAT4 testing can form the basis of plans for their future development, which they themselves can take some control over.

### **Feedback on reports**

Every school in the standardisation received these reports, and where possible feedback was sought to gather teachers' views. Although only 22 schools responded to the survey, the feedback is still worth noting as it is relevant to how these teachers in the Chinese context viewed CAT4.

One theme that came through was how CAT4 could be used to make decisions about their students at the individual level<sup>9</sup> and the breadth of student profiles in their class. They felt CAT4 would help them to:

- “better understand the students, [...] to make more detailed recommendations, so that children can develop”
- “understand the students' multiple intelligences, not only to examine students' IQ, but also their academic achievement.”
- “understand students' different learning circumstances”.

Others mentioned the importance of students' themselves understanding their own profiles:

- “CAT4 can effectively help children more deeply understand themselves, and then get better development. Students should not only focus on the textbooks, given that the development of students' expertise is essential for each child”

One even mentioned the benefit to moving away from rote learning for students: “Use knowledge, rather than rote memory!”.

---

<sup>9</sup> Responses have been translated into English

Another mentioned the importance of measuring students at the beginning of their secondary school journey:

- “If we use CAT4 to test students’ ability on entry to school, teachers will know where students are at the beginning of their studies”.

Some teachers focused on how CAT4 would help them with planning:

- “Unlike other assessments, the CAT4 test focuses on students' long-term development and can help schools to plan their abilities for a long time”.
- “A more comprehensive test of students' cognitive ability, it is a true formative assessment, focusing on long-term development, to help schools’ long-term planning of various abilities of students, [it is] not just a diagnostic assessment”.

When asked what measures they would take after reading their reports, teachers responded that they would “implement a personalised solution” for each child, and they would “individualise teaching plans, drawing on each student’s strength, [and] targeted teaching”. Another wrote:

- “According to the teaching content, based on the requirements of the core literacy, to effectively improve the classroom teaching methods. Strengthen the cultivation of students’ thinking ability, improve students' ability of autonomous learning, and explore interdisciplinary teaching.”

Another teacher wrote about the importance of CAT4 for helping inform future careers:

- “It enables the school to have a more scientific understanding of each student's cognitive level and potential, helping to show the levels and abilities of students for their future career, to help guide the long-term development of each student”.

Interestingly, although CAT4 was administered only to a certain age group of children in this study, there was a call from teachers to adapt further levels of CAT4, to enhance understanding of student potential across other year groups in the whole school:

- “I hope that each grade of students are able to know their strengths and weaknesses, but also for us to understand the different stages of the child's development”
- “I want to learn about the students in different year groups”
- “We should understand students at different levels, in different year groups and involve them – in order to discover similarities and differences.”
- “Enable the school to have a more scientific understanding of each student's cognitive level and potential – to guide all levels and abilities of students in the future career design – effectively to help guide each student's long-term development.”

### **Next steps**

GL Education, together with partners, successfully developed and implemented an adaptation of CAT4 within state schools in China, producing a nationally standardised test of ability.

CAT4’s objectives are to diagnose students’ strengths and weaknesses, and provide advice to guide students in their learning, and teachers in their strategies, planning and decision making. As such, it is important to reflect on the details of how these objectives can be delivered at the teacher and student levels, and how teachers and students can take ownership of them.

Feedback from schools was positive post standardisation. However, a wider study should now be conducted to understand how to start the process of embedding CAT4 within Chinese schools moving forward. What is important is that teachers in China are able to administer

the tests, and start to feel confident in interpreting results, but most importantly, can make use of the data available to them.

There is a danger that post standardisation, the value of these tests and what they can bring can be lost. Indeed, a study by Liu and Feng (2005) found that although teachers were happy to talk about the curriculum reform in kindergarten in China, and self reported their beliefs that teaching ideas, which included respecting children, active learning, and play-based teaching and learning were imperative, most of these ideas were not then put into their teaching practice in reality.

With this in mind, earlier in 2019, GL Education has signed a 10 year partnership with GP Menco, a training partner in Shanghai. The partnership aims to use CAT4 data for evidence-based teacher training in state schools in China, and ensure schools take ownership of administering and interpreting CAT4 data across a much wider school base.

As we move into the next phase of this development of CAT4 in China, we will aim to work with GP Menco closely. It will be an exciting next study to look at any impact that is being made on teachers and students through the use of CAT4 in Chinese schools, keeping in mind the complex needs of such a large education system and comparing how this might vary compared to schools in the UK.

## References

- Cui, Y., Lei, H., & Zhou, W. (2018) Changes in School Curriculum Administration in China. *ECNU Review of Education*, 1(1), 34–7. <https://doi.org/10.30926/ecnuroe2018010103>
- GL Assessment (2012) Technical report: CAT4. GL Assessment: London
- GL Education (2018) Technical Report on the CAT4 Nationwide Trial in China. GL Assessment: London
- Li, H., Wang, C. & Wong, J.M.S. (2011) Early Childhood Curriculum Reform in China, *Chinese Education & Society*, 44:6, 5-23, DOI: 10.2753/ CED1061-1932440601
- National Bureau of Statistics of China (2014), *China Statistical Yearbook*, China Statistics Press, Beijing, [www.stats.gov.cn/tjsj/ndsj/2014/indexeh.htm](http://www.stats.gov.cn/tjsj/ndsj/2014/indexeh.htm) (accessed on July 2015).
- OECD (2016) *Education in China: A snapshot*. OECD: Paris.
- Smith, P., Fernandes, C., & Strand, S. (2003). *Cognitive Abilities Test 3: Technical Manual*. London: nferNelson

Acknowledgements to the following who have been involved in the study and ongoing project of CAT4 in China:

- China Future School Lab (National Institute of Education Sciences)
- Cres Fernandes, Tom Gallacher – Statistics team, GL Assessment
- Jim Kell, Director of International Partnerships
- GP Menco, new partner in China