Dimensions Educatives

Dimensions Test

Towards a full-spectrum universal screening procedure

A tier-based assessment of school achievement, cognitive abilities and emotional development

Eudald Correig Fraga
Dimensions Educatives

Abstract

In this work we present a framework for the development of a full spectrum cognitive, emotional and social adaptation screening procedure. New information technologies, as well as advances in the fields of data science allow for a leap forward on the way we think about psychological and educational test batteries. We propose a shift form the individual, costly and lengthy traditional tests to a new set of flexible and adaptive tests that can gather larger amounts of measurements in very little time and without the need of a psychology trained person. We do this by administering the test in a tiered system of sub-tasks together with a machine learning based data analysis. In this work we also present the first of the tiers of the test and a study its results.

Keywords

Cognitive ability, emotional intelligence, social adaptation, multiple intelligences, test battery, machine learning.

Contents

Introduction

The need for a better monitoring of children's proficiency in the different areas that comprise the basis of their education is very clear. The lack of standard, system-wide, easy to administer tests in our educational system forces professionals to use either individually administered tests [1, 2, 3, 4], which are very time and resource consuming or collective tests, generally cheaper and less time-consuming, but outdated or poorly adapted [5]. Finally, the modern and easy to administer tests that exist nowadays in the Catalan system (i.e. [6]) are very domain specific and thus fail to give a broad image of a child's capabilities.

In this work, we present the Dimensions Test, a set of digitally administered tasks that will attempt to address these issues. The main objectives of our work are:

0.1 Objectives

1. Assess academic achievement and measure and identify all those factors that affect it. Practically this means that we will develop an academic achievement test, together with a cognitive abilities test and completed with tests in non-cognitive abilities and external factors that can affect a children's proficiency at school. We shall shortly explain how we aim to achieve this endeavour.
2. Use the results of the different subtests to further study the reasons causing any observed underachievement. The main causes that we aim to study are the following:
   - The most important subject of study is the link between what we have traditionally known as cognitive abilities and academic achievement, so we are able to pinpoint the deficient abilities that may have caused an academic underachievement.
   - The relationship between skills outside of the traditional definition of cognitive abilities (e.g. emotional intelligence, social abilities, self-recognition, etc) and academic achievement.
   - Social and environmental factors that we know are very important in the academic development of a child (e.g. socioeconomic status).
   - Clinical factors that may affect the learning capabilities of a person.
3. Once these causes are understood, work towards interventions and treatments to be able to effectively address these underlying factors and thus revert any academic underachievement.

Intermission

There are many tests in the market; one may ask why would we endeavour the difficult and perhaps bold task of creating a new one instead of using an already consolidated one. The reasons are the following:

- There are no Catalan translations of the full spectrum achievement and cognitive abilities existing tests. Albeit there are narrow abilities tests in Catalan, current research shows that it's very important to assess children using a full battery test, given that we need to assess children's intellectual profile as a whole in order to correctly understand it [7, 8].
- We need a test that's cheap and easy to administer; given the current education system conditions, the only way of ensuring that any school has access to these tests is to provide them at a very low cost.
- As we mentioned, we want to test what have traditionally been called non-cognitive abilities included in the multiple intelligence theory [9] and cognitive abilities that are not currently in the CHC approach, and therefore, on the big current test batteries (e.g. rationality [10]).
- It's clear that we need to include all the preceding points while shortening the test duration and eliminating the need for psychology professionals for most of the subjects that take the test. As will be explained shortly, we shall do it by using a tiered system that includes dynamic assessment and limit theory.

0.2 Specific Needs

We have been mentioning learning difficulties, special educational needs or other similar expressions without exactly specifying what we mean by them. We are going to address this in this section.

We begin by considering a classroom in a typical school, either public, private or semi-private. If we assume that the students of this class have been chosen at random, we can approximate the performance of the students in any cognitive or achievement task as a Gaussian distribution [12], like the one in Figure 1.

The resources that schools receive, both financially and in personal resources are insufficient, and prevent them to provide an education adapted to the needs of every student [12]. Each individual teacher and school as a whole faces the decision of focusing the limited resources in some students. Naturally, what most schools do is to focus on the average-performance students, given that, as we can see in the figure, is where there are a bigger number of children.

That is, schools and teachers place their limited resources where they can benefit most students [3]. The situation, then, is that the students both at the lower and higher tails often find that their educational needs are not fully covered.

[1]This assumption can be somewhat relaxed, given that in a class with "the smartest" kids, for instance, their distribution will still be normal. This would not be the case if we pick a class where we put together on purpose the bottom and lower range of students form another, bigger distribution. But we can safely assume that this is almost never the case

[2]Research shows that this is exactly true in reading ability tasks [11], but no research exists on other areas. Nevertheless, we believe that it stands to reason that the same applies for all cognitive abilities.

[3]A mention on a great number of resources placed on the very low-performing students is also needed. We will see shortly, though, that these resources are often not very effective

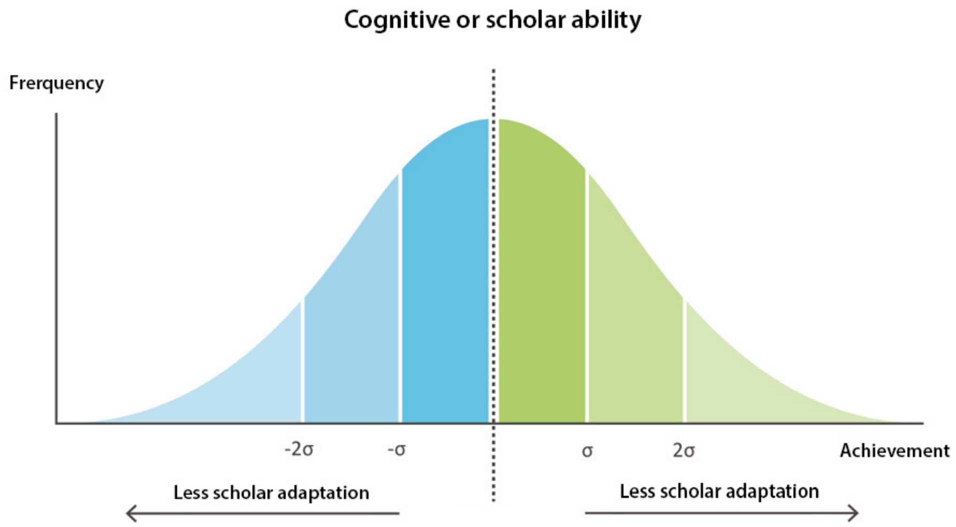

Figure 1: Normal distribution with sigma values and their percentages

Furthermore, a student can (and will) be in the average range in one subject or cognitive area but in the lower or higher tail in another. This means that in one subject he or she will do fine, but in another might have problems. Therefore, a student is usually not either completely well-attended, or poorly attended, but a mixture of the two.

This is then what we mean by special needs: some extra or different attention that a student might need in one or more subjects in their education and development. For example, a child with reading difficulties will need to devote more time to reading, and will need a slower and more individualized path towards reading proficiency. On the other hand, a student who is brilliant in maths, will need deeper and more advanced knowledge in this subject in order to be able to achieve a performance in accordance to her abilities [4]

In the following sections we will see, then, how the "Dimensions Test" (DT) identifies all these needs and their causes (either cognitive, emotional or social) and advises those responsible of their education in how to understand and act on these needs.

0.3 Universal screening procedure

Detection of a learning difficulty or a special learning need of a child at school is a rather haphazard process, given that there are no automatic detection protocols and it depends mainly on the teachers shrewdness and intuition. Even when a special need is detected, the process of pinpointing the exact problem and drawing a solution is very lengthy. This is obviously a huge drawback, given that many children don't have their needs properly detected and others receive their intervention much later than what would be advisable.

It's for these reasons that we have devised the DT as a universal screening procedure, by which

[4]The percentages of children needing extra attention vary widely upon many factors and shall be treated in the Annex.

all children in the school are screened at the beginning of the course in order to give the teachers an idea of which are the specific needs and the potential of every child in every class. This may seem an overwhelming endeavour, but we believe we have devised a procedure to make it work in a reasonable amount of time and resources.

0.3.1 Tiered system

This procedure is based on a tiered test system. That is, the complete Dimensions Test consists of three tiers, with different tests and evaluations in each. We shall describe the different sub-tests in each tier in the next section, but we give here the basic outline of our system. The first tier consists of a collective tests consisting of 8 or 9 short sub-tests (around 3 minutes each) which evaluate academic achievement, some very specific executive functions and cognitive abilities and emotional ability and social adaptation. With this set of sub-tests we can draw a broad profile of the cognitive and social and emotional profile of a child and identify possible risk factors that might affect their education and personal development.

This part of the test is administered collectively by a teacher or support staff to a classroom with access to a digital device such as a computer or a tablet and with good internet access.

Once the test is finished, the algorithm suggests new tasks (belonging to the second tier of the DT) that will be conducted to focus on the cognitive abilities that we know affect the school achievement test in which the first tier has detected a learning difficulty or a special need. We have based ourselves in the current research on human intelligence and in the psychometric studies of past tests, as well as on our research, to find the relationship between the detected needs in school achievement and personal development and the underlying cognitive abilities that affect them. This research has allowed us to create an artificial intelligence system that draws cognitive profiles and links school achievement with underlying cognitive ability.

The tests of the second tier are on-line based aswell and individually administered (again, without the need of a psychologist), in such a way that allows us to evaluate cognitive abilities in their most isolated and pure form as well as emotional or adaptive factors affecting the school achievement. Once this second set of sub-tasks has been administered, we will have a complete image of the cognitive, emotional and social profile of the child; this image will be more "pixelated" in the areas where we find average or good performance, and much more focused in the areas where we detect a special need.

Along with such profile, the test results offer an automatic report on the potentialities and the possible special needs of the child in our educational system, in order to provide the school with the tools to offer the most tailored education possible. This might include interventions on learning or adaptation difficulties, recommendations for deeper emphasis in some subjects or other procedures, both inside or outside the school that can help the child. On the other hand, if any of the difficulties, both learning or adaptation difficulties are considered severe, the school will be advised to administer the third tier of the test.

This third tier consists of a semi-automated intervention lasting between 12 and 24 weeks. If this intervention is not successful in improving the achievement to levels close to the mean, we set up a session with a psychology professional, who will further study the causes for the difficulties and will devise more intensive intervention plans in order to address them. The DT provides with extra sub-tests, questionnaires and other materials to assess these severe difficulties, and, if needed, to diagnose some psychological or learning condition.

0.3.2 Profile examples

To illustrate further how the DT operates, let us give a couple of examples of possible cognitive profiles that the test identifies:

Profile 1

Imagine a child who scores above the average on all tests except the reading fluency one, where the child has scored in the lower range. Given the theoretical framework that the test is based on, it will interpret that there might be a difficulty in decoding speed thus in phonological awareness (i.e. a typical dyslexic). Therefore, the second tier will provide with a set of sub-tests on phonological awareness, as well as other visual and auditory perception areas that might also contribute to this case of reading difficulty. Furthermore, written and oral comprehension tasks might also be nededed. Suppose the results of these test put the decoding speed of the child between the 10th and the 30th percentile, with problems in phonological awareness but not on other areas. This means that we are dealing with a child with a mild reading impairment. The test will recommend for a decoding and phonological awareness intervention in order to give this child a boost and help him or his education.

We can also have the case where the result of the second tier sets the child below the 10th percentile. This is considered a severe dyslexia, and therefore the kid has a high risk of school underachievement. In this case, then, the analysis of the second tier will advise for an individualized evaluation with a professional. This psychology professional with use the material that deems necessary to pinpoint the affection and to intervene on it if possible.

Profile 2

Another possible profile would be that of a child who scores on the lower range on all school achievement tests, very low on the emotional and adaptation tests but on the average on the processing speed test. Any of the current school achievement tests will predict that this child's performance is very low at school and therefore needs of an individualized attention program or even a personalized curriculum. But what the DT will identify is that, instead of a cognitive difficulty, there might be an emotional or adaptation problem that affects very strongly the school performance of the child.

The second tier, then, will administer sub-tests to assess the possible dis-adaptations of the child and questionnaires that will inquire on the external factors that might affect his or her education and personal development. If these affectations or factors are considered mild, a report on how to better adapt this child in the class and the school will be administered. Otherwise, if they are considered severe, again the Dimensions system will advise for the need of a session with a psychologist.

1 Theoretical Background

The aim of this chapter is to outline the scientific background that supports the different sections and tasks of our test, as well as the relationship between them.

The Dimensions Test is based primarily on the Cattell–Horn–Carroll (CHC) theory of human intelligence, which has based both research and tests in intelligence for more than 50 years. But modern research is leaning towards a broader view upon cognitive abilities and general intelligence, tending to assert that other areas in human cognition are equally important in the educational and personal development of children and adults. Many theories have risen to build upon or expand CHC [9, 13, 14], all of them with strengths and shortcomings.

Our experience in working in interventions on learning difficulties and on re-educations for children has taught us that, probably, the cognitive areas outside of CHC that affect children the most are the emotional and social ones[1]. Indeed, emotional difficulties (that might be caused by a huge amount of variables) has a very important effect on how a child experiences school, relation with others and develops as a person. It's for this reason that we have integrated aspects of the multiple intelligence theory, that we believe can will this emotional and social intelligence gap left by CHC.

Furthermore, we will also be integrating some areas of rationality theory, that, as we will shortly explain, fills a very important gap within CHC theory that most tests overlook.

1.1 Cognitive theories Basis

1.1.1 CHC theory

The Cattell–Horn–Carroll theory (CHC), is the most widely accepted theory on the structure of human cognitive abilities. The Cattell–Horn–Carroll theory is an integration of two previously established theoretical models of intelligence: the Gf-Gc theory of fluid and crystallized intelligence from 1941 [15] and 1965 [16] respectively, and Carroll's Three-Stratum theory (1993) [17], a hierarchical, three-stratum model of intelligence. Due to substantial similarities between the two theories they were successfully amalgamated to form the current Cattell–Horn–Carroll theory.

The three stratum model, as we can see in 1.1, consists on a base stratum made of some 80 so called narrow abilities, which factor to 10 second order (or broad) abilities, which, in its turn, factor to one single entity, called the "g" factor. This theory is still being constructed and evolved, so there is still some debate on the broad (stratum II) abilities, and the narrow (stratum I) abilities. [2]

[1]There is some debate on whether these areas should be called cognitive areas, but in this work we use the term "cognititve" in the broadest possible interpretation

[2]We refer the reader to [17, 18, 7] for a comprehensive explanation of this theory.

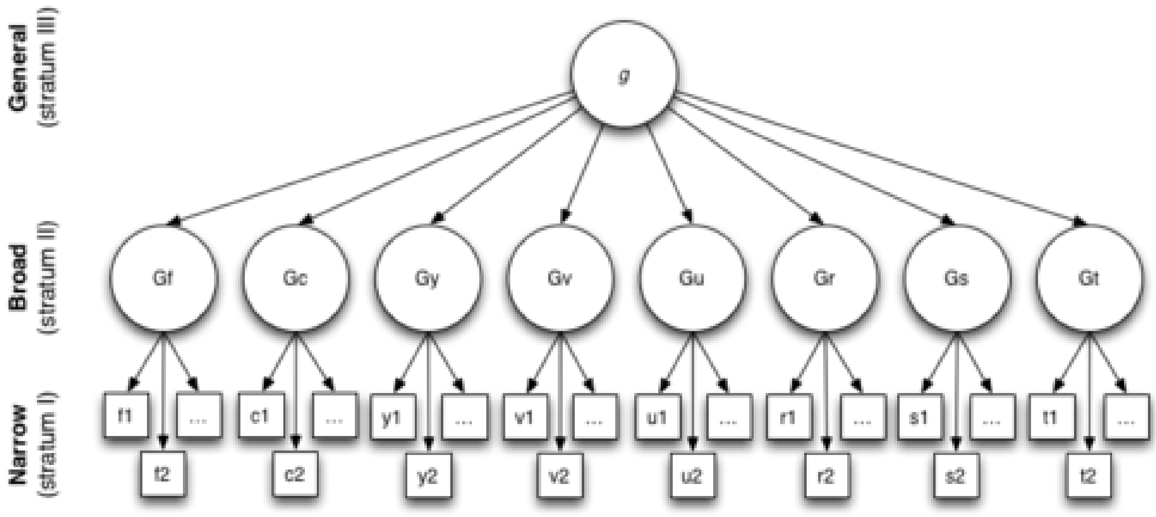

Figure 1.1: Carroll's Three Stratum Model of Human Intelligence: General intelligence (g) fluid intelligence (Gf), crystallized intelligence (Gc), general memory and learning (Gy), broad visual perception (Gv), broad auditory perception (Gu), broad retrieval ability (Gr), broad cognitive speediness (Gs), and processing speed (Gt).

The three-stratum model is well supported by evidence because it is based on Carroll's comprehensive psychometric meta-analysis of 461 diverse datasets meeting specific criteria. Carroll is careful to emphasize that abilities in each stratum merely reflect their levels of generality in governing a range of cognitive abilities; intermediate strata may exist between the three he identified.

But CHC is not free from critique, the most notorious one is clearly exposed in J. Davidson and I. Kemp review [19] on current theories of intelligence:

> [...] the psychometric approach and its models seem to have at least two shortcomings. The first has to do with our criterion that models be based on relevant assumptions. It is not clear that psychometric theories meet this requirement; they rest on the supposition that analyses of scores, from tests taken once, reveal the true structure of intelligence. Test taking occupies a relatively small part of most people's lives and it does not necessarily reflect their intelligent behavior in daily problem-solving situations. Even though the scores are moderately predictive of school achievement and work success [...], they fall short of capturing many aspects of what is considered intelligence.

Since we aim to find *all* factors that affect school achievement, and not just scores that are moderately predictive, we will include tests based in other theories that complement CHC and that can help us get a better picture of the cognitive profile of the testee. In the next section we will outline these theories.

1.1.2 Cross Battery Assessment

Before going to other theories of human cognition, we need to understand further how test evaluate the enormous amount of cognitive abilities that comprise CHC.

With a few exceptions ([20, 21, 3]), current test batteries lay its scientific foundations on the CHC theory of human cognitive ability. An analysis of the current intelligence and cognitive abilities tests is outside the scope of this work, but recent reviews have concluded that all test batteries fall short in some of the broad cognitive abilities [22].

To address this issue, in 1990 D. Flanagan, S. Ortiz and K. McGrew [23] devised the so-called Cross Battery Assessment (XBA), whereby they created a framework to allows psychologists use information from multiple test batteries to help gaining a fuller picture of the cognitive abilities of the test subject than could be ascertained through the use of only one battery assessment.

Albeit a great improvement form previous single-battery assessments, it's rather clear that this is just a patch on the shortcomings of previous tests, and that the creation of a single test battery would be both theoretically and practically more sound than this approach. The problem with a full-encompassing battery is that, if we would follow the current test administering methods, it would take a huge amount of time to test a single subject.

Facing this apparently bleak scenario, we have devised a test structure that we believe will solve the test-length problem and adapt the XBA approach of testing every aspect necessary at the same time as overcoming the impracticality of having to use more that one test battery and, at the same time, reducing the time needed to administer the test.

1.1.3 Multiple Intelligence Theory

The theory of multiple intelligences, proposed by Howard Gardner in 1983 is a theory of intelligence that differentiates it into eight specific modalities, rather than seeing intelligence as dominated by a the general factor "g". The abilities that he suggested are musical–rhythmic, visual–spatial, verbal–linguistic, logical–mathematical, bodily–kinesthetic, interpersonal, intrapersonal, and naturalistic [9].

Exploring the definitions of these different abilities, we can see that the visual-spatial, the verbal-linguistic and the logical-mathematical are already encompassed by CHC theory, but the other ones are not, and hence are not included in most of the cognitive abilities tests.

Extensive research [24, 25] asserts that all these different intelligences play a major role in the development of a child, both in the scholar context and outside. Furthermore, there have been many examples of educative systems that have adopted MI theory thesis in their educational program, generally with great success [26, 27, 28].

In light of this, we include tests of these other intelligences in our battery to provide tools and resources for educative centers to better implement MI theory in their program and to better equip families to booster the education of their children in areas that may be neglected in the academic curriculum.

Again, we refer the reader to [9, 29] for a more comprehensive explanation of this theory.

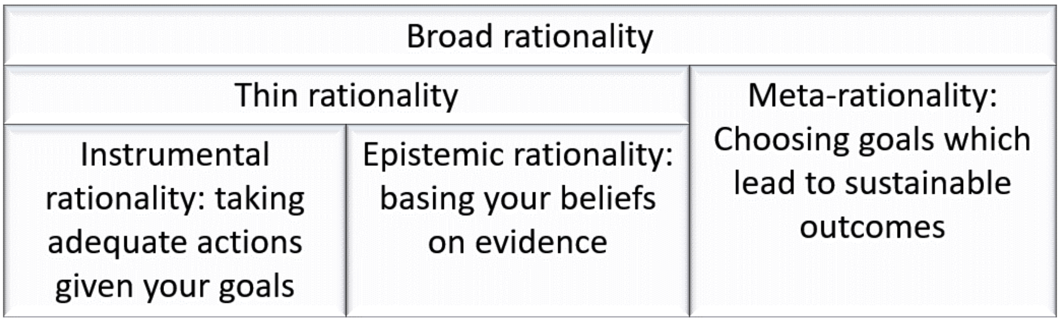

Figure 1.2: Outline of the different aspects of rationality theory by Stanovich.

1.1.4 Rationality theory

According to Keith Stanovich, developer of this theory, Rationality is the "the optimization of the individual's goal fulfillment". In other words, being rational means acting to achieve one's own life goals using the best means possible. Essentially, this encompasses the notion of consciously choosing the best strategy to surmount any type of problem.

In his presentation of this theory, he argues that this type of intelligence is separate for the other "modalities" of intelligence, both in the theoretical and in the factor analysis sense. For a more detailed explanation, please see [13].

Traditional IQ tests do not test for this ability [30, 31], so Stanovich provides ways to test it [32]. It is clear that this type of intelligence influences academic achievement and child development, so we adapt these new tests in our test battery.

1.2 Inner Test Logic

1.2.1 Boston Process Approach

It is obvious that we cannot test all we said we were going to test in a reasonable amount of time by following conventional screening methods. This is where the Boston Process Approach (BPA) enters. BPA is an approach to neuropsychological assessment whereby the focus is put on understanding the underlying processes used by the testée in order to undertake a particular exercise. This approach acknowledges the fact that two different people can get the same total score in a test even though they have used different processes and have completely different cognitive abilities, and, therefore, different strengths and weaknesses. For more information, see [33, 34]

The way we implement this idea in our test battery is to organize it in a tier structure. [3] The first tier consists of a screening where we test for the broad categories of the elements that we have stated above. A computerized automatic evaluation of the results in these tests is performed in order to analyze the errors and weaknesses of the evaluée. For this purpose, we do not only analyze the total score of each test, but also the relationship among tests in order to get the best possible cognitive picture.

[3] The complete structure will be explained in the next chapter.

Based on these results, we provide a new set of individualized and specific tests that deepen only on the narrow cognitive abilities that cause the weaknesses and provide a complete picture of the possible causes of underachievement.

1.2.2 High risk subjects

After the two first tiers, we pick those subjects with higher risk factors, both related to learning difficulties, and to emotional problems, and offer them with a third tier of tests. These tests include an individualized session with a psychology trained professional who is going to assess the clinical and learning status of the testée. In this assessment, the psychologist is going to be able to chose, depending on the results of the previous Tiers, the tests or other types of evaluations better suited for each case. Therefore, risk subjects are going to receive a complete evaluation of their learning and emotional abilities, as well as their clinical profile, both by the tests themselves and also by a professional. Steps on how to proceed and corrective measures are going to be suggested or implemented by this professional if needed.

1.2.3 Dynamic assessment

As part of this third tier, the psychology professional can implement a period of training and post-evaluation (that can take weeks or months) in order to dynamically assess the learning aptitudes and the emotional and clinical evolutions of the subjects, and to better evaluate the real condition and sources of the risk elements.

According to the last Diagnostic and Statistical Manual of Mental Disorders (DSM-V), dyslexia can only be diagnosed after a period of at least 6 months of intervention. If the patient is resistant to this intervention, then we can diagnose dyslexia [35]. In the DSM this is only explicit for Dyslexia, but it stands to reason that this idea can be extended to other learning difficulties. The third Tier in the Dimensions Test adapts seamlessly to this procedure, given that this dynamic assessment will include, if needed, the intervention required to address any deficiency and at the same time work toward a complete diagnostic.

2 Test Structure

The Dimensions Test is going to be organized in three Tiers.

2.1 First Tier

The first Tier will consist in an academic achievement test, and it will be possible to administer either collectively as a paper-based test, or individually as an on-line test. This test is designed to be able to stand alone as an academic achievement tests, which can be used by schools and families to monitor their children's performance in school and to test whether there are any risk factors that can affect children's maturity.

As we will shortly see, the test consist of four purely academic achievement sub-tests plus two cognitive and two adaptation evaluations that will help us pinpoint the possible deficiencies in the other four tests and will give us the necessary information to construct a second set of cognitive sub-tests adapted to the testée. As we will see, this new set will form the second Tier.

The sub-tests that comprise the first Tier are the following:

General Reading

This is a reading fluency test. It measures the ability to read and understand short sentences fast. The test is designed in such a way that the underlying and most important construct that measures is decoding speed. The main reason for poor reading skills in children is slow decoding, so we make this our priority measure [36, 37]. Furthermore, the test is organized in for sections, where other factors are also tested. These are vocabulary knowledge, morphology, semantic field and syntax. Following a statistic treatment of the results in the overall test and its relationship with the results in the subsections will allow us to acquire a reliable indicator of overall reading fluency, as well as hints at the specific low achievement sub-construct that might affect it [38].

Working memory

Working memory (WM) involves holding information in mind and mentally working with it (or said differently, working with information no longer perceptually present [39, 40]. There are two types of WM distinguished by content—verbal WM and nonverbal (visual-spatial) WM. Both are included in our test (except for the first cicle of school, where there is only the visual-spatial part). WM is critical for making sense of anything that unfolds over time, for that always requires holding in mind what happened earlier and relating that to what comes later. Research shows that WM is a central element in the reasoning processing path and therefore a very important element in the cognitive profile [41, 42].

Visual processing Speed

In this test we measure the ability to make quick decisions based on visual input and simple conceptual processes. Processing speed is a basic element in any intelligence theory, and will affect overall performance in most of the other cognitive abilities [43]. Furthermore, most of our tests are time-limited, which means that the processing speed will have even a greater impact in those. In measuring it separately we aim to statistically subtract its influence on the other cognitive processes in order to be able to have a purer measure of these other processes. The reason that we do it through visual stimulus is because it's only way to measure this dimension in a collective setting.

Mathematical fluency

A mathematical fluency test is the equivalent of decoding but in mathematics. The test consist of a series of easy arithmetic tasks that consist of addition, subtraction, multiplications and divisions that must be completed as fast and as correctly as possible. With this test we measure the level of automatic mathematical processing, which is the basis of more complex mathematical abilities [44].

Long term memory

Long term memory is the ability to store information over long periods of time, from around two minutes (depending on the theory) to a lifetime. Any learning process is obviously based on the long term storing of information so there is no need to emphasize the need of an accurate measurement of this ability to assess the learning capabilities of an infant. For further information on long term memory and its exact relationship with other types of memories can be found in [45].

Note: The knowledgeable reader might be missing at this point a measurement on Short-term memory. This is not an oversight by our part, but instead the incapability of finding a strategy to measure short-term memory in a collective setting. It's for this reason that we leave it to the second Tier.

Reasoning

This test measures the ability to reason with concepts and to identify logical relationships and patterns and to predicting the outcome of these patterns. In other words, reasoning is the ability to draw conclusions from the merging and evaluation of both previous acquired information and new and unknown input. The human being is goal driven, so these conclusions are paramount in the problem-solving and decision-making processes throughout life [46].

Calculus

In this test we measure more complex mathematical processes, whereby the test subject must work with basic mathematical arithmetic tasks, but in such a way that non-automatic mathematical processing is needed. This test is only performed on children above 9-10 years (5th degree in catalan primary school). It aims at identifying two types of profiles: one with normal or high automatic mathematical fluency but low performance on more complex mathematical abilities. Such a profile might be an example of either a child that hasn't acquired mathematical rules for external reasons or that has a reasoning difficulty. In this second case the outcome of the next sub-test will reaffirm this suspicion and will point us to the path of the underlying abilities that may be faulty. The second profile that we aim to identify is that of a very strong mathematical processing that can help us in identifying children with a great potentiality in logical and mathematical performance.

Sociability and emotional intelligence

In this test we measure the ability of the subject to adapt to its surrounding social environment and to understand and control its own emotional state. This is a very important and often overlook subject when it comes to understand children behaviour and achievement at school. Emotional or social ability deficiencies can affect scholar achievement the same way a poor decoding ability can. Therefore, we need to measure it in order to give a complete child development image.

2.2 Automated report

Once the test has been administered to the whole school and the results entered in our system an automatized report is generated. Here we give an outline of the statistical analysis and the report:

We compute the percentile of each of the tasks both at the universal (all our data) level and the classroom level. We believe that the two measures are important, given that they often do not coincide. To get an image to why they might be useful, imagine a class with a very high level in mathematics. A student might score at the normal level in the universal scale, but still be at the lowest percentile range at the classroom level. This kid will probably not need an intervention in mathematics, but he or she might be needing some extra attention from the professors in order to follow the class pace.

With this results we aim to provide information to whether the student might need a specialized intervention or some extra attention by the school or the family in order to have their specific educational needs met.

But this is not all: as most tasks (except the memory tasks and the emotional part) are time constrained, the execution speed of the student is a central factor in most of the sub-task scores. But we want to measure the ability in each measured dimension regardless of the execution speed. This is why we have developed the Dimensions Method for statistical analysis. This method has four steps:

- We predict the "pure" execution speed from the results from all the sub-tasks using an algorithm.
- We construct a model, based on the results of all our test administrations, in which we model the results in each task as a function of all other tasks *and* the previously calculated pure execution speed.

- We predict the results of each sub-task of each student using this mode.
- We measure the difference between the predicted and real values. If these differences are bigger than some predefined threshold we identify it as a special need.

Therefore, with this analysis we try to identify those areas where the achievement is inferior (or clearly superior) to the expected value according to the overall cognitive profile, and are therefore those areas where a specific intervention might be most needed and effective.

2.3 Second Tier

The goal of the second Tier is to provide fine-grained results to identify the cognitive abilities that are affecting aspects of the performance of the first tier.

This test will be only digital and the results will be provided automatically at the end of the test.

The cognitive abilities that we measure are in synchrony with the theories described earlier and include the following: firstly, the 7 CHC broad categories, namely: fluid intelligence, crystallized intelligence, visual processing, short-term memory, long-term memory, auditory processing. Furthermore, we include tests in intrapersonal intelligence (that is, recognizing and managing emotions both in oneself and in others, and managing relationships), musical intelligence and natural intelligence (perception of the world around oneself) that belong to the multiple intelligence theory. Finally, we include a test in rationality, which we develop from rationality theory. At last, if needed, further clinical tests will be administered.

The relationship between the performance on the achievement test and the underlying cognitive abilities is a complex one. Albeit there are many studies on the subject [17, 5, 1] that we rely upon, the aim of these tests is also to improve the knowledge on this relations by studying the behaviour of the testées. To do this, we will construct automatic statistical measurements and studies of the results, that will improve our knowledge in this area, as well as the tests, in a feedback relationship that will make the test a learning entity that will improve constantly. The basis for these relationships will be outlined in the next section.

The automatic analysis of the results of the second tier will provide a detailed report on the cognitive picture of the child, as well as the possible risk factors that it might have detected. If the risk factors are important, the test will recommend the child to undergo an individualized session with a psychology professional, which comprises the third and last Tier of the Dimensions Test.

Regardless of the need of this third Tier, the results of the second Tier will also recommend specific activities to address the possible difficulties that it has detected and to diminish the risk factor.

Therefore, the third Tier will be an individualized session with a psychology professional. This Tier is designed for those who have shown great factor risks (less than fifth percentile in some tests), as well as those who could be having important emotional instabilities or clinical problems. In this cases we deem the intervention of a psychology professional absolutely necessary.

2.4 Third Tier

This section of the test is reserved to the cases that require the presence of a trained psychologist in order to evaluate and diagnose, if necessary, any severe difficulties or psychological adaptation issues. Thus, two types of profiles will need to take this phase of the test: first, children who perform on the lowest percentile [1] in any of the school achievement or cognitive ability task. A child with such a severe difficulty will definitely need of an intensive intervention, and our experience show that probably will need of psychological support and guidance in order to have successfully follow the intervention and overcome the difficulty. This is only possible through a trained psychologist that can diagnose any difficulty and design and supervise the intervention to overcome it.

The second profile is that of a child who might or might not have a school achievement difficulty, but who, on the emotional and clinical tests hints at a possible adaptation problem. The diagnosis and possible reeducation of these children need to be done and supervised by a psychologist.

The third Tier will consist then on any test or evaluation that the psychologist might find necessary, as well as the guidelines for a possible intervention or reeducation. After this evaluation has ended, the Dimensions Test will have done its job and a psychology center will take over in order to ensure a correct adaptation of the child.

[1] The exact percentile range of the "lowest percentile" depends upon many variables and will be treated in the Annex.

3 Psychometric analysis

A lot of information can be extracted from the data that the Dimensions Test produces. As we shall see, the broadness of the test allows for a richer analysis and construction of a cognitive profiling than do most of the conventional tests.

In this section we present first a descriptive analysis of the validation phase of the test, later we move to correlations between subtasks and then we move to a factor analysis. Finally, we present an exploratory clustering analysis, where we try to identify common cognitive profiles and their particularities.

All analysis have been carried out using the R software version 3.4.3.

Note: We keep the emotional part of the test outside of the psychometric analysis for the time being, given that it doesn't readily transform to a measurable scale end thus will be treated separately.

3.1 Descriptive statistics

The validation phase took place during the months of April to June 2017 in 7 Catalan schools (not disclosed for anonymity). These schools were spread throughout the Catalan geography balancing both the size of the villages and cities where they stood, the type of school (public, semi-private, private) and the socioeconomic status of their alumni.

The Catalan primary school lasts for 6 years, from 5-6 to 11-12 years of age. We validated the test in these 6 courses. In table 3.1 we give values for number of children tested, mean and standard deviation for each sub-task and for each of the six courses in which we validated the test. Note that the scores are corrected by subtracting the correct answers minus the incorrect ones multiplied by a factor that depends on the sub-task and that ranges from 0.2 to 0.5. Also note that the tasks are different between courses, so direct comparison is not possible.

We also include a table for the calculus test 3.2, keeping in mind that only applies to the fifth and sixth courses.

3.2 Normality tests

For this and the next sections we standardize each sub-task within its course (subtracting the course mean and dividing by the course standard deviation) before pooling scores across courses, since the raw tasks differ from grade to grade and are not otherwise comparable. Rather than imputing missing values, the correlation matrix uses pairwise-complete observations and the factor and clustering analyses use complete cases only ($n = 3{,}284$ student sessions with all six non-calculus sub-tasks present).

		Primary school grades					
		1st	2nd	3rd	4th	5th	6th
Subtask	n =	331	412	365	367	266	275
Reading	mean	14.65	21.42	15.36	19.77	23.22	24.86
	sd	7.38	7.41	5.83	6.32	7.64	8.52
WM	mean	4.25	4.61	12.52	13.83	17.42	18.40
	sd	2.33	2.55	3.76	4.01	6.04	5.46
VPS	mean	20.31	26.66	30.00	32.78	36.13	37.87
	sd	8.51	8.00	5.47	5.64	5.37	6.62
MF	mean	21.39	29.36	35.88	42.66	48.32	55.20
	sd	10.28	11.92	9.63	11.21	13.17	14.78
LTM	mean	4.19	4.79	10.96	12.60	13.81	14.17
	sd	2.33	2.43	4.16	4.24	5.11	5.04
Reasoning	mean	10.37	12.78	14.91	17.16	18.24	18.32
	sd	4.66	4.48	4.36	3.88	3.97	4.82

Table 3.1: Descriptive statistics of all cognitive sub-tasks on all primary school grades.

		Primary school grades	
		5th	6th
Subtask	n =	266	275
Calculus	mean	21.59	25.10
	sd	7.22	8.02

Table 3.2: Descriptive statistics of the calculus sub-task for the fifth and sixth primary school grades.

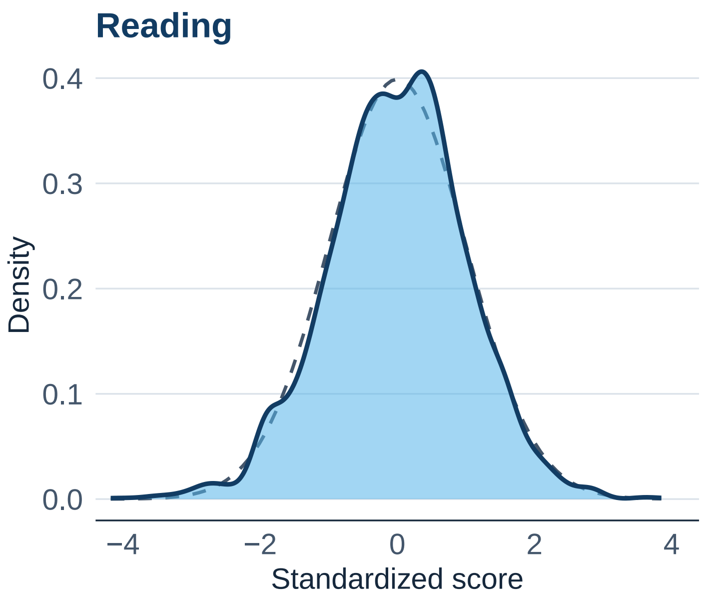

Figure 3.1: Distribution of the centered and scaled results for all courses of the reading sub-task (in blue), where we have drawn a normal distribution with mean 0 and standard deviation 1 to compare (dashed line).

Already in '92, Shaywitz showed that reading ability follows a normal distribution [11]. Therefore, the general results of a reading ability tests should also form a normal distribution. In picture 3.1 we can see that indeed our results (in blue) follow very closely a normal distribution (black dashed line). This is the desired result of the test and we believe it proves that it's a powerful and reliable reading test.

There is no research to be found on the theoretical or experimental distributions of the other areas that we measure. Nevertheless, it stands to reason to hypothesize that all cognitive abilities might follow a normal distribution, at least at the theoretical level.

We have studied our results, shown in 3.2, and the outcome is that they do not follow a normal distribution. This can be due to several reasons: first, the true distribution is not normal, although we don't think this is the case. Second, the theoretical distribution is normal but, due to educational biases, the achievement distribution is not normal. Third, it might be that everything is normally distributed, but our tasks fail to capture it. We believe the answer is a mixture of the second and third reasons, but more research is needed, both at the theoretical and testing levels.

Nevertheless, the distributions are well-spread and smooth and clearly allow for the construction of a useful achievement scale in each one of the sub-tasks.

Once we have explored our data we can start delving in the inner structure analysis:

3.3 Inner structure

We compute the inner validity of the test using Cronbach's alpha and Guttman's Lambda 6 [47, 48], over the six non-calculus sub-tasks and $n = 3{,}284$ complete student sessions from

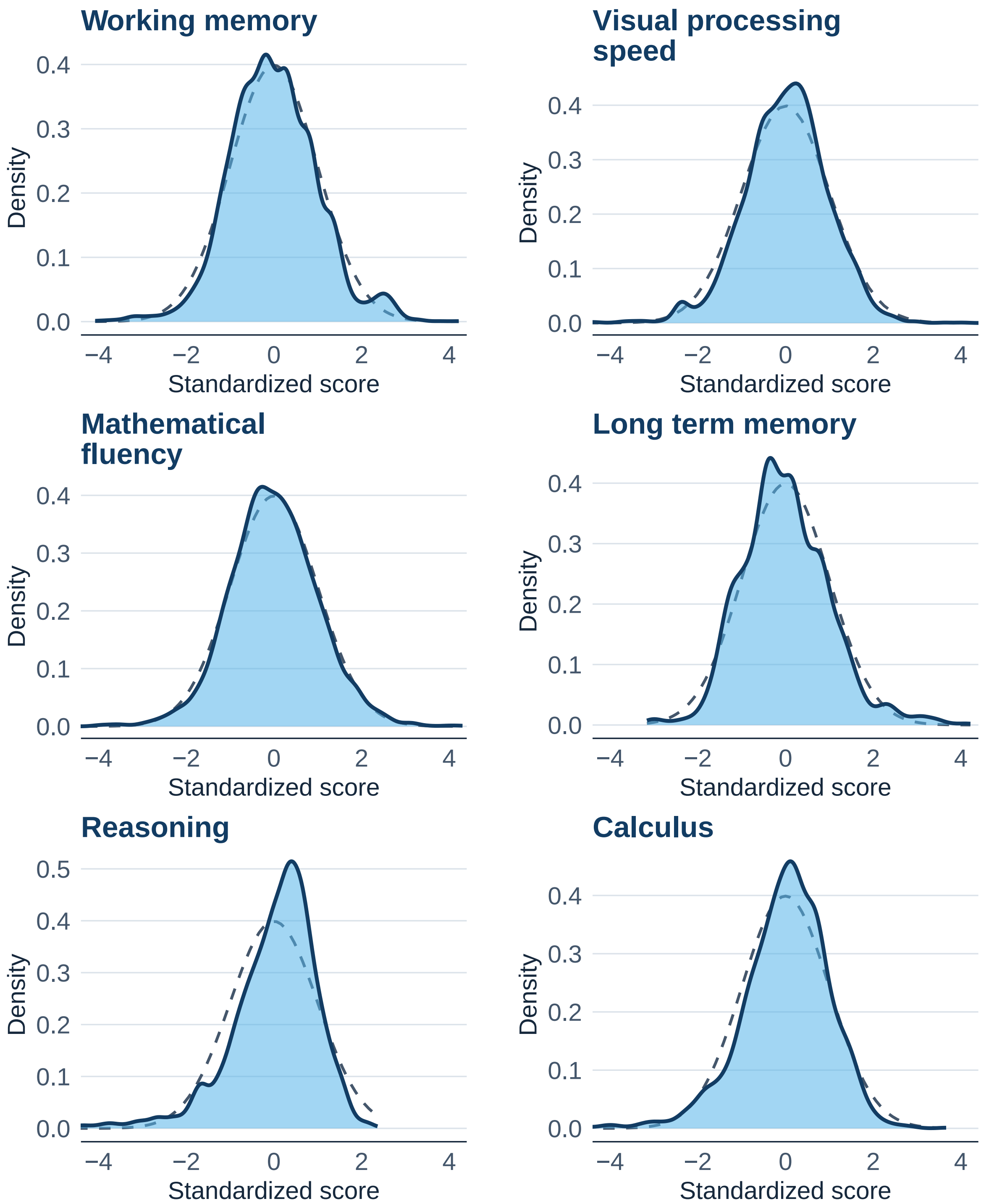

Figure 3.2: We can see that the other results diverge from a normal distribution but they are sufficiently spread to allow for good sensitivity. Note that the calculus sample is smaller ($n \approx 2{,}058$) than the other six sub-tasks ($n \approx 3{,}290$ each), since calculus is only tested from course 5 onward.

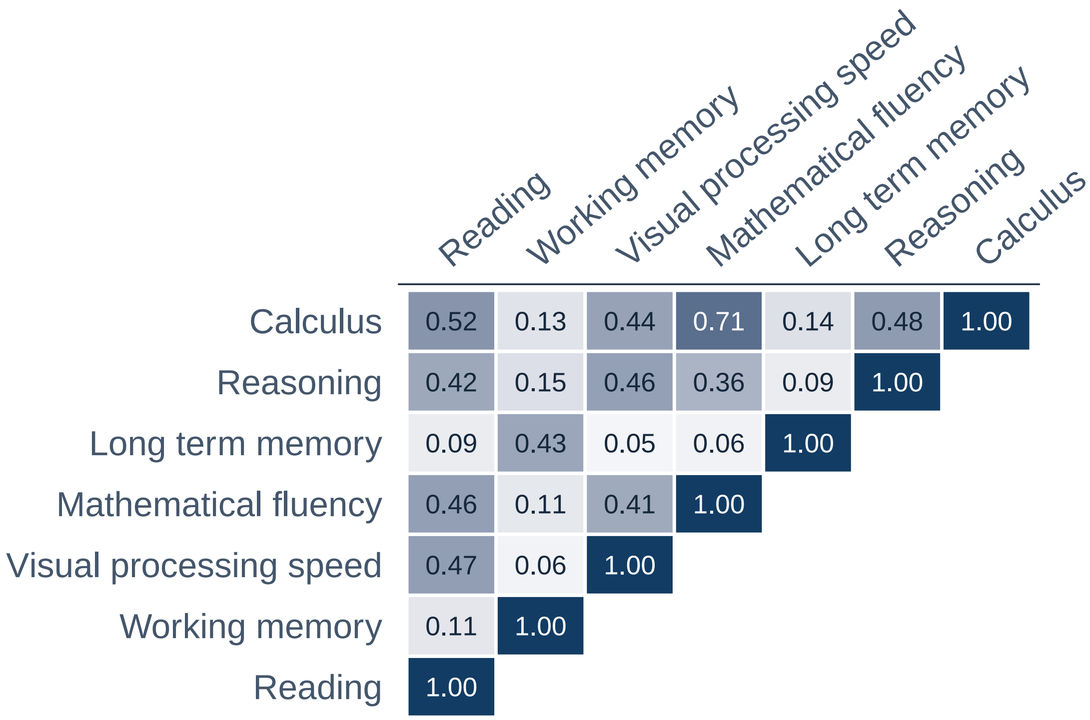

Figure 3.3: Heatmap of the correlations between the sub-tests. Note that the calculus correlations have only been computed among students in course 5 and above, since calculus is not part of the test in earlier grades.

current production data. Their values are 0.66 and 0.67 respectively: a moderate, rather than excellent, inner consistency. This is expected – and arguably desirable – given that the sub-tasks were deliberately designed to probe different cognitive dimensions rather than a single underlying ability; a very high reliability would instead suggest the sub-tasks were redundant with one another.

Furthermore, we explore the correlation matrix of the results to gain insight on the observed relations between the different variables.

In this case we see that the two memory subtasks (Working memory and Long term memory) remain fairly independent from the rest: they correlate strongly with each other (0.43) but weakly with everything else (0.05 to 0.15). The other five subtasks – Reading, Visual processing speed, Mathematical fluency, Reasoning and Calculus – are moderately intercorrelated (roughly 0.36 to 0.52), with Mathematical fluency and Calculus standing out as the most related pair (0.71), unsurprisingly given how much numerical fluency underlies calculus performance. This partially confirms our design intent: the subtasks were devised to measure different cognitive

Loadings			
Cognitive dimension	PC1	PC2	PC3
Reading	-0.501	0.121	-0.232
Working memory	-0.203	-0.670	0.008
Visual processing speed	-0.487	0.197	0.258
Mathematical fluency	-0.466	0.130	-0.674
Long term memory	-0.165	-0.690	-0.037
Reasoning	-0.475	0.070	0.651

Table 3.3: Loadings of the six cognitive dimensions on the first three principal components (unrotated).

dimensions, and memory indeed stands apart, but the language/speed/math/reasoning tasks now share more common variance than the smaller 2017 validation sample suggested.

In the next two sections we continue exploring this idea in factor and a clustering analysis.

3.4 Factor analysis

In this part we will proceed to do an exploratory principal component analysis of the results in the different sub-areas of the test. This analysis checks for latent variables that lead to correlations between the tests that we can detect. It is important to note that the different sub-areas of this test have been chosen as to expand the widest possible range of cognitive and emotional measurable dimensions and therefore we expect to have little correlation between the variables and so weak latent variables.

Table 3.3 shows the loadings of the six non-calculus dimensions on the first three components of an (unrotated) principal component analysis, computed on the same complete-case sample used for the clustering biplots below ($n = 3{,}284$). Unlike the 2017 validation, we no longer apply a varimax rotation: with a much larger sample the unrotated components already separate cleanly by construct, and rotating them would only relabel the same structure. The first component (39% of the variance) loads with similar weight on Reading, Visual processing speed, Mathematical fluency and Reasoning, and is best read as a general non-memory performance axis. The second component (23% of the variance) is dominated by the two memory dimensions (Working memory: -0.670, Long term memory: -0.690), confirming memory as its own factor. The third component (11% of the variance) contrasts Mathematical fluency (-0.674) against Reasoning (0.651), a fluency-versus-reasoning axis that the smaller 2017 sample did not resolve.

Furthermore, table 3.4 shows that the first three components now account for 73% of the variance across the six dimensions – noticeably more structure than the 2017 validation sample suggested. Memory still stands apart as its own axis (PC2), but the remaining five sub-tasks share a sizeable general-performance component (PC1) and a further, more specific fluency-versus-reasoning contrast (PC3).

	PC1	PC2	PC3	PC4	PC5	PC6
SS loadings	2.351	1.365	0.662	0.589	0.522	0.512
Proportion variance	0.392	0.228	0.110	0.098	0.087	0.085
Cumulative variance	0.392	0.619	0.730	0.828	0.915	1.000

Table 3.4: Variance captured by each principal component.

Therefore, this analysis confirms that memory remains a distinct cognitive dimension, while also revealing more shared variance across the other sub-tasks than the original validation implied – consistent with the moderate (rather than very high) reliability reported above.

We will study this further in the following section:

3.5 Clustering analysis [1]

We will base the study of the main achievement profiles on an unsupervised clustering analysis in which we will explore the predominant characteristics of the different test results. The clustering analysis has been done with a k-means clustering algorithm with k=5. The number of clusters has been chosen following the elbow rule [49]. We present the results of these clusters in two principal components (PC) charts: in the first (3.4) we show the first and second (PC), and in the second (3.5) we show the first and third PC.

As we can see in 3.4 and 3.5, the data allows us to draw very clear clusters of the different profiles.

We believe that this type of information can be of enormous interest to the teaching community, given that it can allow teachers and professors to use strategies designed to have maximum effect on certain cognitive profiles.

For further study, we also think that it may be very interesting to study of the clustering movements along the ten courses now covered by the test, from primary through secondary education. This will be an excellent tool to gauge the shortcomings of the educational system. That is, we believe it's very important to identify certain profiles with low achievement in one or more areas that are in a standstill and tend to not be able to glide to "safer" areas of the achievement chart. These profiles are the clear candidates of more attention and particularly of specialized tools and methods to help them inch to a normal educational progress. Exploring and defining these profiles in the most accurate possible manner is of paramount importance in order to succeed in this endeavour.

[1] This is a somewhat more technical section, it can be skipped without losing the overall threat. For more information on the techniques used we recommend to visit the Wikipedia pages on Clustering Analysis and Principal Component Analysis and study their cited works.

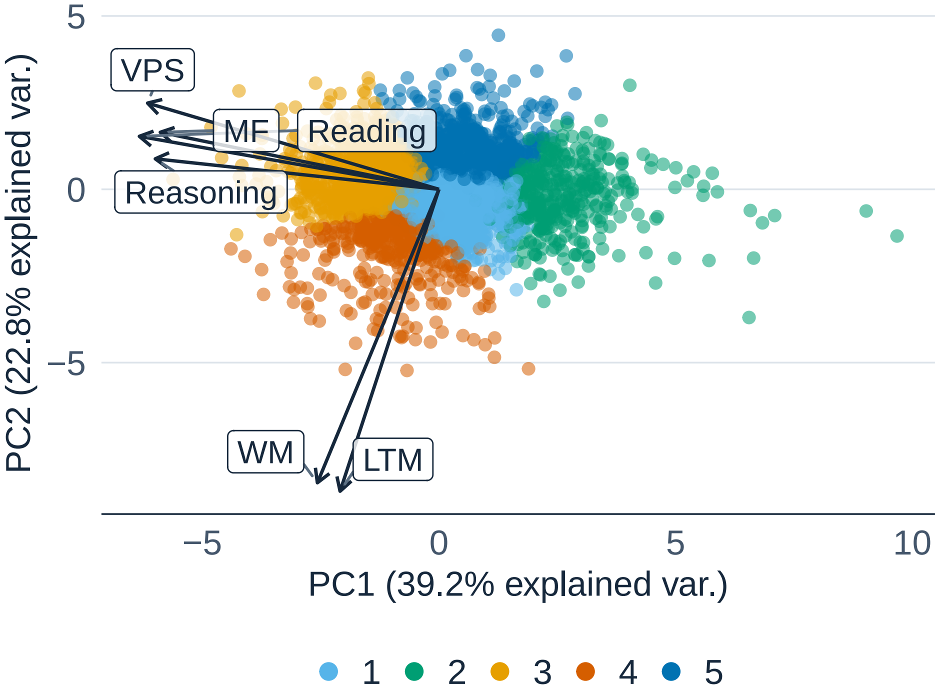

Figure 3.4: Unsupervised clustering of $n = 3{,}284$ current production student sessions into 5 clusters. Arrow labels use the same abbreviations as table 3.1: WM (working memory), VPS (visual processing speed), MF (mathematical fluency), LTM (long term memory).

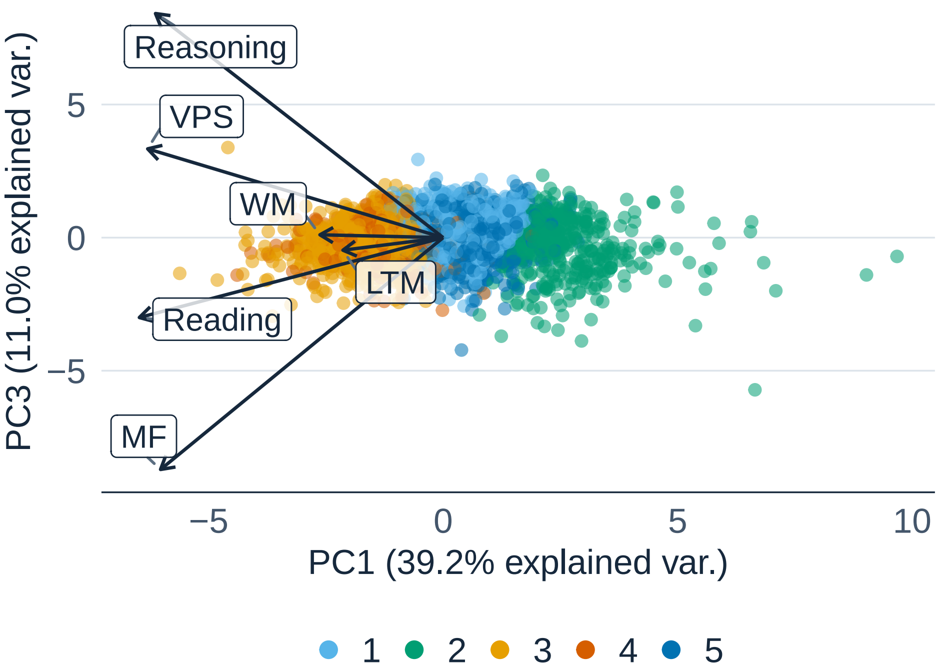

Figure 3.5: Unsupervised clustering of $n = 3{,}284$ current production student sessions into 5 clusters. Arrow labels use the same abbreviations as table 3.1: WM (working memory), VPS (visual processing speed), MF (mathematical fluency), LTM (long term memory).

3.6 Conclusion

The results of the psychometric analysis are very satisfactory since they mainly validate our hypothesis. First,the distributions of the different sub-tasks, albeit not all normal, show that the tasks have enough sensitivity to correctly construct an achievement scale.

Second, the inner reliability measures are moderate (Cronbach's $\alpha = 0.66$, Guttman's $\lambda_6 = 0.67$) rather than excellent. This is expected, and arguably desirable: a single high-reliability score would suggest the sub-tasks are redundant measures of one ability, working against the test's purpose of profiling several distinct cognitive dimensions.

Third, the principal component analysis has shown that memory (working and long-term) forms its own, largely independent dimension, while the remaining five sub-tasks – reading, visual processing speed, mathematical fluency, reasoning and calculus – share a moderate general-performance component alongside a more specific fluency-versus-reasoning contrast. This is a more nuanced picture than the original validation's near-total independence between sub-tasks, but it still supports the broader hypothesis that the test captures more than one cognitive construct.

Finally, in the clustering analysis we provide a glimpse of the power of the overall test to draw cognitive profiles and to help researchers, teachers and families understand the needs and strengths of each student.

Summary

In this work, we introduced a framework for developing a new kind of educational strategies based on the student's needs and strengths. The set of tools we develop within this framework provide the educational community are based on the current scientific ideas and tailored to meet the needs of the Catalan educational system.

The first Tier of the Dimensions Test, which is what we have presented in this work, is only the first of these tools, but, given the results presented, we believe provides a very strong stepping stone to keep developing and implementing the rest of the tools outlined in the introduction of this work.

Future tools in the Dimensions context

In this last section, we present the tools that will be developed after the first Tier of the Dimensions Test:

1. **Tier II:** As explained in previous sections, the first Tier of the Dimensions Test is "just" a preliminary school achievement test that gives us a broad picture of the underlying cognitive abilities, emotional intelligence and social adaptation of the student. Now it's time to develop Tier II, where we delve on the measurement of the fundamental dimensions that comprise human cognition. This is definitely a daunting enterprise and shall be the main purpose of the future work at Dimensions.

2. **Tier III:** For the third Tier of the test, we need to develop methods and protocols for our psychologists to perform the final emotional and social adaptation evaluation and to develop guidelines for any required intervention.

3. **Intervention** Development of these mentioned interventions, both at the cognitive and emotional and adaptation levels.

Bibliography

[1] D. Wechsler, "Wechsler adult intelligence scale–fourth edition (wais–iv)," 2014.

[2] A. S. Kaufman, S. E. Raiford, and D. L. Coalson, *Intelligent testing with the WISC-V*. John Wiley & Sons, 2015.

[3] R. L. Lemon and C. A. Waehler, "A test of stability and construct validity of the black racial identity scale, form b (rias-b) and the white racial identity attitude scale (wrias).," *Measurement and Evaluation in counseling and development*, 1996.

[4] J. Portellano, R. Martínez-Arias, and L. Zumárraga, "enfen (evaluación neuropsicológica de las funciones ejecutivas en niños)," *Madrid: TEA ediciones*, 2009.

[5] K. S. McGrew and R. Woodcock, "Technical manual," *Woodcock-Johnson III*, 2001.

[6] F. C. Vega, *PROLEC: Batería de evaluación de los procesos lectores de los niños de Educación Primaria*. 1998.

[7] W. Schneider and K. McGrew, "The cattell-horn-carroll model of intelligence," *Contemporary intellectual assessment: Theories, tests, and*, no. 3rd, pp. 99–144, 2012.

[8] J. B. Carroll, "The three-stratum theory of cognitive abilities," *Contemporary intellectual assessment: Theories, tests, and*, no. 3rd, pp. 883–889, 2012.

[9] H. Gardner, *Frames of mind: The theory of multiple intelligences*. Basic books, 2011.

[10] K. E. Stanovich and R. F. West, "Advancing the rationality debate," *Behavioral and brain sciences*, vol. 23, no. 05, pp. 701–717, 2000.

[11] S. E. Shaywitz, M. D. Escobar, B. A. Shaywitz, J. M. Fletcher, and R. Makuch, "Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability," *New England Journal of Medicine*, vol. 326, no. 3, pp. 145–150, 1992.

[12] B. Albaigés and G. F. Esteban, "L? èxit educatiu a catalunya.: Indicadors dels sistemes educatius," *Escola catalana*, vol. 47, no. 478, pp. 20–21, 2012.

[13] K. Stanovich, *Rationality and the reflective mind*. Oxford University Press, 2011.

[14] J. P. Das, J. A. Naglieri, and J. R. Kirby, *Assessment of cognitive processes: The PASS theory of intelligence.* Allyn & Bacon, 1994.

[15] R. B. Cattell, "Some theoretical issues in adult intelligence testing," *Psychological Bulletin*, vol. 38, no. 592, p. 10, 1941.

[16] J. L. Horn and R. B. Cattell, "Refinement and test of the theory of fluid and crystallized general intelligences.," *Journal of educational psychology*, vol. 57, no. 5, p. 253, 1966.

[17] J. B. Carroll, *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press, 1993.

[18] J. O. Willis, R. Dumont, and A. S. Kaufman, “Factor-analytic models of intelligence,” *The Cambridge handbook of intelligence*, pp. 39–57, 2011.

[19] J. E. Davidson and C. Downing, “Contemporary models of intelligence,” *Handbook of intelligence*, pp. 34–49, 2000.

[20] S. Urbina, “Tests of intelligence,” *The Cambridge handbook of intelligence*, pp. 20–38, 2011.

[21] J. A. Naglieri and J. P. Das, “Cognitive assessment system.,” 1997.

[22] D. Flanagan, V. Alfonso, and S. Ortiz, “The cross-battery assessment approach: An overview, historical perspective, and current directions,” *Contemporary intellectual assessment: Theories, tests, and*, no. 3rd, pp. 459–483, 2012.

[23] D. P. Flanagan and K. S. McGrew, “A cross-battery approach to assessing and interpreting cognitive abilities: Narrowing the gap between practice and cognitive science.,” 1997.

[24] J.-Q. Chen, S. Moran, and H. Gardner, *Multiple intelligences around the world*. John Wiley & Sons, 2009.

[25] A. Craft, H. Gardner, and G. Claxton, *Creativity, wisdom, and Trusteeship: exploring the role of education*. Corwin Press, 2007.

[26] M. L. Kornhaber, E. G. Fierros, and S. A. Veenema, *Multiple intelligences: Best ideas from research and practice*. Allyn & Bacon, 2004.

[27] E. Barrington*, “Teaching to student diversity in higher education: How multiple intelligence theory can help,” *Teaching in Higher Education*, vol. 9, no. 4, pp. 421–434, 2004.

[28] J. Wagmeister and B. Shifrin, “Thinking differently, learning differently.,” *Educational leadership*, vol. 58, no. 3, pp. 45–48, 2000.

[29] H. E. Gardner, *Intelligence reframed: Multiple intelligences for the 21st century*. Hachette UK, 2000.

[30] K. E. Stanovich, *What intelligence tests miss: The psychology of rational thought*. Yale University Press, 2009.

[31] M. E. Toplak, K. Stanovich, and R. West, “Intelligence and rationality,” *Cambridge handbook of intelligence,*, pp. 784–826, 2011.

[32] K. E. Stanovich, R. F. West, and M. E. Toplak, *The rationality quotient: Toward a test of rational thinking*. MIT Press, 2016.

[33] E. Kaplan, “The process approach to neuropsychological assessment,” *Aphasiology*, vol. 2, no. 3-4, pp. 309–311, 1988.

[34] W. P. Milberg, N. Hebben, E. Kaplan, I. Grant, and K. Adams, “The boston process approach to neuropsychological assessment,” *Neuropsychological assessment and neuropsychiatric disorders*, pp. 42–65, 2009.

[35] A. P. Association *et al.*, *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[36] R. K. Wagner and J. K. Torgesen, "The nature of phonological processing and its causal role in the acquisition of reading skills.," *Psychological bulletin*, vol. 101, no. 2, p. 192, 1987.

[37] S. A. Stahl and B. A. Murray, "Defining phonological awareness and its relationship to early reading.," *Journal of educational Psychology*, vol. 86, no. 2, p. 221, 1994.

[38] M. A. Just and P. A. Carpenter, "A theory of reading: From eye fixations to comprehension.," *Psychological review*, vol. 87, no. 4, p. 329, 1980.

[39] A. D. Baddeley and G. J. Hitch, "Developments in the concept of working memory.," *Neuropsychology*, vol. 8, no. 4, p. 485, 1994.

[40] E. E. Smith and J. Jonides, "Storage and executive processes in the frontal lobes," *Science*, vol. 283, no. 5408, pp. 1657–1661, 1999.

[41] P. C. Kyllonen and R. E. Christal, "Reasoning ability is (little more than) working-memory capacity?!," *Intelligence*, vol. 14, no. 4, pp. 389–433, 1990.

[42] A. Diamond, "Executive functions," *Annual review of psychology*, vol. 64, pp. 135–168, 2013.

[43] A. R. Jensen, *Clocking the mind: Mental chronometry and individual differences*. Elsevier, 2006.

[44] W. Fias, V. Menon, and D. Szucs, "Multiple components of developmental dyscalculia," *Trends in neuroscience and education*, vol. 2, no. 2, pp. 43–47, 2013.

[45] N. Cowan, "What are the differences between long-term, short-term, and working memory?," *Progress in brain research*, vol. 169, pp. 323–338, 2008.

[46] J. P. Leighton, "Defining and describing reason," *The nature of reasoning*, pp. 3–11, 2004.

[47] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *psychometrika*, vol. 16, no. 3, pp. 297–334, 1951.

[48] L. Guttman, "A basis for scaling qualitative data," *American sociological review*, vol. 9, no. 2, pp. 139–150, 1944.

[49] R. L. Thorndike, "Who belongs in the family?," *Psychometrika*, vol. 18, pp. 267–276, Dec 1953.