ADAPTIVE MATRIGMA

Technical Manual

Second Edition

HELENE HOPPE REVALD
Director of Psychometrics, Occupational Psychologist

RALUCA IOANA CHIFOR
Psychometrician, Occupational Psychologist

DENNIS HVASS
Psychometrician, Occupational Psychologist

Assessio

Zero Talent Waste.

Table of Contents

1. Introduction and theoretical background

This manual describes the background and development of Adaptive Matrigma, Version 1.0. Adaptive Matrigma is a further development of the non-verbal general mental ability (GMA) test, Matrigma. Because of the connection between these tools, it is recommended that readers familiarize themselves with the technical manual for Matrigma (Mabon & Sjöberg, 2016) before reading this manual. The Matrigma technical manual presents, for example, the theoretical background, the importance of GMA in general, the measurement of this ability in particular, the relevance of matrices as a format, and the connection with outcomes such as performance at work. These areas are also relevant to Adaptive Matrigma. Note also that this manual presumes the reader has a basic knowledge of psychometrics and is well acquainted with Classical Test Theory. For those who need a refresher on these subjects, the books Scale Construction and Psychometrics for Social and Personality Psychology (Farr, 2011) and (in Swedish) Arbetspsykologisk testning (Mabon, 2014) are recommended.

This manual begins with a definition of general mental ability (GMA), which is what Adaptive Matrigma measures. After that, we explain why GMA is important and what can be derived about a person based on their test results. We also instruct the user on how to apply and interpret the test correctly. Finally, we detail how Adaptive Matrigma was created and describe all the research we have done to ensure its quality.

General Mental Ability (GMA)

In the early twentieth century, Charles Spearman, introduced the concept of a general intelligence factor - commonly referred to as the "g factor." Spearman's research revealed that individuals who perform well in one area of cognitive testing are likely to perform well in others, suggesting the presence of a broad, underlying mental capability that influences performance across diverse intellectual tasks. However, Spearman also acknowledged the presence of specific abilities, known as "s factors". These are unique to particular tasks or domains and operate alongside the general factor. Later, Raymond Cattell refined the understanding of intelligence by distinguishing between two broad components: fluid and crystallized intelligence. Fluid intelligence encompasses the capacity to solve novel problems, reason abstractly, and adapt to new situations independent of prior knowledge. In contrast, crystallized intelligence represents the knowledge and skills acquired through experience and education, such as vocabulary and factual knowledge. Interestingly, fluid intelligence tends to decline with age whereas crystallized intelligence tends to remain stable or increase throughout the lifespan (Ryan et al., 2000).

The relevance of GMA in a business context

In the business context, intelligence tests are widely recognized for their ability to predict job performance and workplace success. General Mental Ability has consistently been shown to be one of the strongest predictors of training outcomes, problem-solving capacity, and overall job performance across a wide range of roles and industries (Schmidt et al., 2016). By assessing intellectual capabilities such as reasoning, comprehension, and problem-solving, organizations are better able to identify candidates who are likely to learn quickly, adapt to new challenges, and make effective decisions. As a result, intelligence testing has become an integral tool in personnel selection, talent development, and succession planning, helping companies build high-performing and adaptable workforces.

Characteristics of Adaptive Matrigma

Just like Classic Matrigma, Adaptive Matrigma is a non-verbal test with simple geometric figures. Each question displays a 3x3 matrix with 8 two-dimensional figures and a question mark in the bottom right corner. The figures are ordered inside the matrix based on one or more rules. Candidates must figure out what those rules are and replace the question mark with the missing figure. There are six answer options to choose from.

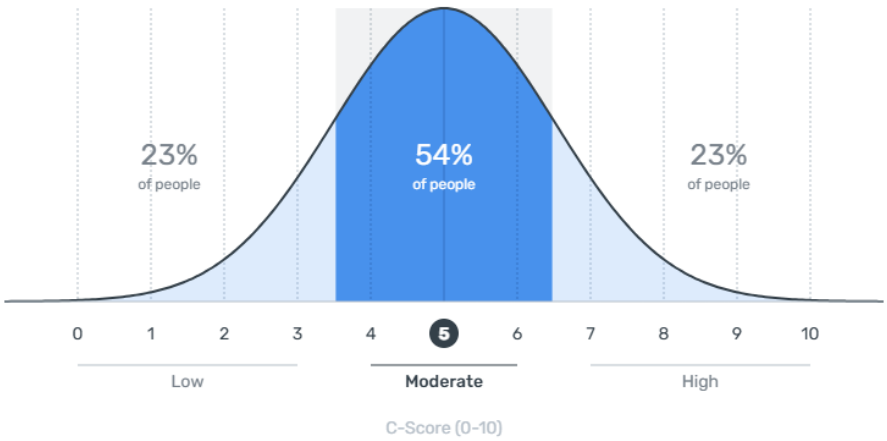
Matrices are commonly used to assess fluid intelligence. Perhaps the most famous test of this kind is Raven's Progressive Matrices (Raven, 1947, 1960, 1998). One reason for choosing this format is that it does not involve language. Therefore, people with a reading disability or who aren't taking the test in their native language will not be disadvantaged. Moreover, solving matrices does not require any specialized knowledge. This means that candidates have equal chances to demonstrate their abilities irrespective of their education level.

As the name suggests, Adaptive Matrigma is a dynamic test which adapts to each candidate. That is, not all questions are equally difficult. People with higher GMA will answer more difficult questions correctly than people with lower GMA. When a candidate starts the test, they will first receive a less difficult question. Based on whether they answer wrongly or correctly, the test will adjust the difficulty of the next questions. Progressively, the difficulty of the test will be calibrated for that specific candidate's ability level. The goal is to find the maximum level for which the candidate gives consistently good answers. This will provide an estimation of their GMA. The advantage of adaptive tests is that they reach an accurate result with less questions. In practice, this means a shorter assessment time for the candidate. In addition, it prevents frustration caused by too many difficult questions or boredom caused by too many easy.

2. Scale definitions and interpretations

Adaptive Matrigma only measures a single construct, General Mental Ability (GMA), without tapping into specific mental abilities. All questions have the same format and are thus counted as one scale. The speed of answering does not impact the final score. The results are displayed as C-scores, which range from 0 to 10. They have a mean of 5 and a standard deviation of 2. Below is a figure showcasing how to interpret those.

Figure 1. Distribution of C-scores and their interpretation for Adaptive Matrigma.



About 23% of people have a C-score between 0 and 3, which indicates a low GMA. An equal percentage score between 7 and 10, thus demonstrating a high GMA. Most people (54%), however, have a C-score between 4 and 6. This indicates a moderate level of GMA.

Low scores
Candidates with low scores tend to struggle with complex problem-solving, requiring additional support. They may have a slower learning curve, finding it more difficult than others to adapt to new or unexpected situations.

Moderate scores
Candidates with moderate scores solve moderately complex tasks with ease and only occasional guidance. They adapt reasonably well to routine tasks and learn new skills at a moderate pace. Whether they need more or less help solving problems depends on how familiar they are with that genre of problems.

High scores
Candidates with high scores excel in complex problem-solving and demonstrate great analytical thinking. They adapt quickly and effectively to new or challenging situations, and rapidly acquire new skills, displaying a strong aptitude for continuous learning.

On the candidate feedback page, the lowest and highest categories are split into C-scores 0-1 and 2-3 and C-scores 7-8 and 9-10, respectively, covering 4% and 19% at each end of the scale. This division is made to give more nuanced scores and interpretations, thus increasing the likelihood of candidates recognizing and accepting their scores.

3. Instructions for use

Areas of use

Adaptive Matrigma has been developed for assessment in the context of personnel selection. The qualities measured are universally important and impact job performance in all professions, therefore making Adaptive Matrigma applicable for any position and for all industries and businesses. Adaptive Matrigma may be used for professions at all levels and in all lines of work, preferably as a first step in the selection process. Adaptive Matrigma can also be used in a development context (ideally combined with other measures, such as personality assessments) to support career coaching, team building initiatives, etc. The format of delivery (web-based and unsupervised administration) and the efficiency of Adaptive Matrigma makes it highly suitable for screening a larger number of candidates. However, there is no inherent barrier for using Adaptive Matrigma on a smaller number of candidates.

Administration and scoring

Adaptive Matrigma can be administered online via our assessment platform. The respondent completes the items shown on screen and the web system computes the raw scores, converts the raw scores into standardized scores, generates results, and provides standardized feedback reports. The test can be completed remotely or on-site with supervision. Adaptive Matrigma always starts with a tour where the test person gets to complete a total of three practice items before starting the actual test. It should be noted that these tasks are not being scored and are separate from the actual items in the test. To take the test, candidates must receive an invitation link. Platform users may send this link via e-mail together with a message. The e-mail template can be edited if needed. It is strongly recommended that the e-mail to the candidate includes information regarding:

- The purpose of testing.
- What type of test Adaptive Matrigma is and why it is being used in the present context.
- How Adaptive Matrigma will be administered and what is required for completing the test.
- What mode of device is to be used. This is especially important if the respondents are given access to complete Adaptive Matrigma using mobile devices.
- How the results will be used and saved, by whom, and for how long.
- If feedback will be provided; if so, when will it be distributed, what format will it be in (standardized on screen, personal feedback face-to-face, over the phone), and what will it contain.
- Contact details for the test administrator.

More information about the rights and obligations of test distributors, test administrators, and candidates are to be found in international guidelines for testing (e.g., www.intestcom.org, www.efpa.eu/professional-development, www.iso.org/standard/56436.html) and is often provided by national psychologists' associations.

Requirements and conditions of testing

To get valid results, certain conditions must be met:

- Reserving 12 minutes to complete Adaptive Matrigma. The time limit is tracked by the web system and is shown on screen. Each item has a time limit of 1 minute. This means that respondents are shown at least 12 items in total. The maximum number of items a candidate can receive is 40. However, it is extremely unlikely the respondent will be able to solve that many questions within the time limit. When the time limit is reached,

the test session will end, and the respondent's responses will be saved. For each item, candidates can choose from six answer options. Only one option is correct.
- Ensuring respondents' basic reading comprehension – although the written instructions aim to be short, simple, and straightforward, they nevertheless require a basic level of reading comprehension.
- Ensuring that the respondent does not suffer from any form of impairment that is likely to have a negative effect on the test result. This may include but is not limited to perceptual, visual, and/or cognitive impairments.
- Ensuring that the Adaptive Matrigma is responded to in a non-distracting environment – public environments, such as internet cafés and public transportation are not suitable for taking Adaptive Matrigma.
- Ensuring that the respondent has access to a personal computer when taking the Adaptive Matrigma. The assessment can be completed using a tablet, smartphone, or similar device but this might sometimes affect visibility.
- Ensuring that the respondent has basic technical skills – the respondents' must for example be able to use a mouse and/or keyboard to complete Adaptive Matrigma. The test administrator should ensure that the technical aspects do not increase the test difficulty for the respondent, as this would have a negative effect on the result.
- Ensuring that the respondent has access to a stable and reliable internet connection for the full duration of the testing in order to ensure a valid result.

Adverse conditions

To ensure they perform at their best, test-takers should try to avoid:
- Taking the test while in a poor mental or physical condition (e.g., not enough food or rest, or while dealing with burnout). Likewise, they should refrain from taking the test shortly after consuming alcohol or any other substances that affect brain functioning.
- Taking the test on a small-screen device such as a smartphone or tablet. The test was optimized for desktop users, and therefore we recommend using a laptop or PC.
- Starting the test at the wrong time or place. For best results, the candidate must be fully focused on the test and free of distractions. We recommend reserving a quiet and comfortable space and making sure that other people (such as family members or roommates) do not enter that space while the candidate is taking the assessment.
- If a candidate is blind or suffers from severe visual impairment, matrices tests such as Adaptive Matrigma are not suitable for them. In those cases, we recommend finding an alternative way to evaluate the candidate's GMA.

Prevention of cheating

Cheating on psychological assessments has always been a risk. Therefore, we do our best to spot it and prevent it. In this section, we explain how we do this.

Because the test is adaptive, candidates receive different questions that are tailored to their level. Moreover, the answer options are ordered differently every time. These measures prevent candidates from memorizing the questions and answers and leaking those to other candidates. All our cognitive tests feature complex images that are hard to read by AI. In addition, the answers to our questions are not searchable on the internet because they don't relate to specific fields of knowledge such as history, literature, or biology. In addition, all questions must be answered within a time limit. Many candidates confess this limit is too short to allow them to use other devices or applications. This means that trying to use a search engine or AI or asking a more "knowledgeable" person to answer the test for them won't necessarily help.

Guidelines for retesting

If necessary, it is possible to re-test candidates without experiencing much of a “learning effect”. A “learning effect” is when people perform significantly better on the test because they have completed it multiple times and therefore learned from their past mistakes. However, Adaptive Matrigma has a large item pool. This means that candidates will get different questions every time they complete the test, thereby reducing the “learning effect”.

Users may opt for supervised or unsupervised re-testing. When re-testing happens with supervision, we recommend conducting the assessment on-site in a designated testing location. While there are ways to proctor remotely, our research shows that candidates prefer being supervised on-site. In a recent study (2026) we asked more than 300 real candidates how they feel about the following proctoring methods: (1) on-site proctoring, (2) remote proctoring with a human proctor watching through their computer camera, (3) remote proctoring with an AI proctor watching through their computer camera. Candidates ranked on-site proctoring as the most comfortable and fair, and the AI-based proctoring as least comfortable and fair. When asked to explain their answers, many candidates expressed concerns related to AI and data privacy or doubted that an AI could do a good job in supervising candidates. They also mentioned feeling more relaxed in the presence of another human and that being able to see what the human proctor does (on-site) reassures them that the supervision protocol is indeed adequate. In case of proctored testing, it is the responsibility of the test administrator to provide the test person with the conditions needed for optimal test completion.

General principles for interpretation and feedback

After the test is completed, the respondent receives a summary of their results. However, this is only applicable if the test administrator enables it. The test administrator can therefore decide whether to provide the respondent with the standard feedback report or not. For the test administrator, the results are always available in a more detailed view. The scores are presented alongside their interpretation, as well as the norm group used. The interpretation is written to be understood by people without a Psychology background. However, for those who want to understand the test and its scoring system in depth, we recommend studying this manual and/or registering for training to become an expert user.

The test administrator may rank respondents based on their scores. This feature is readily available in our platform. Respondents will not see their ranking.

4. Scale construction

Adaptive Matrigma was designed based on Classic Matrigma. The two share the same pool of items (150 in total). They both measure GMA and are grounded in the same theory. They are meant for recruitment purposes and predict job performance with similar accuracy. How they each achieve this differs, however. Classic Matrigma is based on the principles of Classical Test Theory (CTT), while Adaptive Matrigma works according to Item Response Theory (IRT). These two theories differ in how they treat aspects such as validity, reliability, and measurement error. Ultimately, those differences impact how the test is administered and scored, thereby offering a different user experience. In the following paragraphs, we offer some information on CTT and IRT, as well as what differentiates the two. This information is relevant to those who want to understand what makes Adaptive Matrigma different from Classic Matrigma.

Classical Test Theory

Most psychological assessments which exist today were developed according to Classical Test Theory (CTT). The following ideas are fundamental to CTT:

- All items are considered equally important. Therefore, each item answered correctly returns an equal number of points.
- Each item in the test measures the construct in question with a certain amount of error. Those errors are presumed to be random. When more items are summed into a scale, some measurement errors may cancel each other out. However, the final score will most often be an imperfect estimation of the construct in question.
- The accuracy of the final score is called “reliability” and is calculated on group-level. This can be further used to derive accuracy of individual scores. In CTT, the expected error on individual scores is called the “standard error of measurement”.
- When scores are normally distributed (i.e., resembling a Gaussian distribution or “bell curve”), the standard error of measurement is equal across the entire scale This means that the accuracy of an individual score is the same for low, moderate, and high scores.
- Someone’s “true” score (here, GMA level) is estimated by comparing them with relevant others. This group of “others” is also referred to as a norm group and consists of people who have taken the test in similar conditions. The size and characteristics of the norm group influence how scores are calculated and interpreted. Therefore, having a norm group that is large and well-composed is crucial to ensuring a fair assessment.

Item Response Theory

Item Response Theory (IRT) was developed in the 1950’s and started to receive increased attention in the 1970’s. In IRT, test scores aren’t just a sum of correct answers. Three other characteristics are considered, namely the difficulty (b), discrimination (a), and guessing rate (c) of each item (DeMars, 2010).

Difficulty (b) identifies the level at which 50% of those tested are expected to respond correctly to the item. In CTT, difficulty is labelled P and it is defined as the percentage of individuals (in the norm group) who answer the item correctly. Difficulty thus corresponds to the mean value for each item.

Discrimination (a) represents how well an item differentiates between people with different levels of the measured characteristic (here, GMA). Discrimination is therefore a measure of the

effectiveness of an item, with high discrimination being desirable. The point-biserial correlation coefficient is the typical measure of discrimination within CTT. A positive correlation indicates that individuals who responded correctly to an item have a higher aggregate score for the remaining items compared to those who responded incorrectly.

Guessing rate (c) represents the likelihood of responding correctly to an item by chance or through guessing. This parameter is primarily determined by the number of response alternatives: five response alternatives, for example, yields a 0.2 likelihood of guessing correctly. The likelihood of guessing correctly, however, is also influenced by the level of difficulty for each response alternative. For example, if two out of five response alternatives are obviously incorrect, the guessing is de facto among the three remaining alternatives, yielding a 0.33 likelihood of guessing correctly. By correcting for guessing (c), more exact estimates of an individual's level can be determined.

Main differences between classic and adaptive tests

Unlike classic tests which derive scores only from answers to questions, adaptive tests also account for the fact that some questions may be more complex or telling than others. In addition, adaptive tests weigh in the probability that someone answered a question correctly simply by guessing. In classic tests, a person's score is compared against those of relevant others. This group of "others" is also referred to as a norm group. In the case of Classic Matrigma, this means a person's GMA will be considered above average when they score higher than the norm group average, and below average when they score lower than the norm group average.

Norm groups may also be used to determine the difficulty, discrimination, and guessing rate of each question in the test. However, the scores of individual candidates are not compared against the norm group per-se. Rather, a candidate's level is estimated based on the difficulty level of the questions that they answer correctly.

Initial development of Adaptive Matrigma

The development of the Adaptive Matrigma occurred in two main steps. In the first step a pilot version was developed using data from candidates (N=1000) who had participated in one of the five fixed parallel test versions that were produced in 2014 for the classic version of Matrigma. A total of 150 items were analyzed using a full 3-parameter IRT model (the three parameters being difficulty, discrimination, and guessing as described above). The items and the three estimated parameters for each item were implemented in the web-based platform to calculate the results for the pilot version of Adaptive Matrigma. The pilot version was implemented along with candidate instructions and practice tasks primarily based on the corresponding content for the classic version of Matrigma.

To generate even more exact results, a study was conducted using the data that had been collected on the pilot version. The aim of the study was to obtain more refined parameter estimates. This study utilized data from candidates who had taken both Classic and Adaptive Matrigma for selection purposes (N=1,667). Everyone in this sample had been administered one of the five parallel versions of Classic Matrigma and received a C-score indicating their GMA level in relation to the applied norm group of N=4,606. Employing the new parameter estimates for all items, a new theta value (that is, the value expressing a person's ability level in z-score units with a mean of 0 and a standard deviation of 1) was then calculated for each

candidate. The group's (N=1,667) theta values thus formed a distribution which could be linked to the distribution of C-scores that the same candidates had obtained from the classic (norm group-based comparison) version of Matrigma. In this way, the theta value distribution for Adaptive Matrigma could be 'translated' into a C-score distribution. The theta-based C-score obtained through Adaptive Matrigma thus provides information about both the absolute level of GMA and the relative level – in comparison to the Matrigma norm group (see the Matrigma Technical Manual for more information about this norm group).

Overall, the result of implementing the updated parameter estimates is to generate a somewhat higher C-score (approximately 1 C-score) compared to the previous pilot version. Note also that when implementing the new parameter estimates, the instructions and practice tasks were also refined and updated, affecting the candidate's experience. In general, by applying a refined user experience approach, the length of text-based instructions was shortened and simplified, and the practice tasks were made more instructive.

Scoring model in Adaptive Matrigma

To determine an individual's level of GMA the following steps are taken:

1. A random item with known characteristics for the parameters of difficulty (b), discrimination (a), and guessing (c) is administered to the respondent. The first item, item 1, to be administered is always a randomly selected item whose difficulty level corresponds to a C-score of approximately 1.
2. The respondent's correct or incorrect response to the item is registered.
3. If the response to the first item was correct, the second item will be a randomly selected item from among those corresponding to approximately a C-score level of 3, a somewhat more difficult item than the previous. However, if the response was incorrect for item 1, item 2 will be a randomly selected item from among those with a difficulty level corresponding to somewhat lower than a C-score of 3. This means that even if the respondent gives an incorrect response to item 1, the next item will be somewhat more difficult – although not as difficult as it would have been if the response to item 1 had been correct.
4. The selection of item 3 is based on the same premises as the previous item except that the randomly selected item following a correct response will correspond to a difficulty level of approximately a C-score of 5, a somewhat greater difficulty level. An incorrect response to item 2, however, is followed by a randomly selected item that corresponds to a considerably lower difficulty level. A respondent can thus deviate from the "warm-up phase" earlier than another respondent.
5. This is because the initially selected items are on the low side in terms of difficulty level. Consideration is given to the responses to the previous items, but the selection of the next item is on the low side in terms of difficulty level. The items that are administered in this warm-up phase can therefore be considered adaptive, but to a somewhat limited extent. This assigning of relatively lower difficulty decreases successively as the matching gradually improves along with further items being administered.
6. When the 'warm-up' phase is completed, the respondent is assigned an initial theta (q) value called the 'prior.' The prior is decisive in the next step of this iterative process.
7. In the next step, item selection is based on previous responses and the values of the three parameters for a larger group of individuals with the purpose of "finding" the respondent's "true" q value. The difference between the respondent's q value and the likelihood that the person will respond correctly to the next item is registered. The sum of these standardized differences is added to q, with negative differences decreasing the q value and positive differences increasing it. The best estimation of a respondent's

GMA level is then considered to be reached when an item response minimally alters the respondent's q value.

8. This iterative process continues for 10 minutes. This is enough to estimate the person's cognitive ability. During the remaining 2 minutes, the respondent receives questions corresponding to their estimated level. This helps to further calibrate the final score.

5. Validity

Face validity

Face validity concerns the extent to which test users and test takers perceive the questionnaire and results as relevant, comprehensive, and reflective of reality. Face validity is thus about whether a test comes across as credible to the ones using the test, which is important to ensure that they are sufficiently motivated to participate in the test and accept the conclusions drawn from it. Face validity is also about the extent that test results resemble how a person perceives themselves or is perceived by others.

Given the nature of Adaptive Matrigma measuring GMA with purely figure-based items, it might not come across as the most valid nor fair assessment from the perspective of the candidate. In other words, candidates might not understand the inferential leap from using figures to measure GMA, which in turn predicts job performance across different roles. However, several steps can be taken to increase the perceived value and fairness of Matrigma. These steps can be summarized into five BRIEF recommendations:

- **Be nice**: Treat all candidates with respect and appreciate the effort they make in completing the assessment. Acknowledge if candidates find the assessment difficult or struggle to understand the purpose of it.
- **Reduce uncertainty**: Communicate clearly and support candidates to reduce potential stress or anxiety. Make sure to outline (or provide, in the case of proctored testing) optimal conditions for testing as outlined in the section on instructions for use.
- **Inform**: Provide the candidates with detailed information on what the assessment is, what it measures, and how and why it is used. This increases the likelihood of candidates accepting the results and the conclusions drawn from them.
- **Explain**: Explain to the candidates why they were or were not hired. If Matrigma is used as one piece of information, make sure to emphasize that the final hiring decision was not made solely based on the GMA score.
- **Feedback**: Ask candidates for feedback on your recruitment process. This is the best way to get valuable information as to how the process or communication around it can be improved.

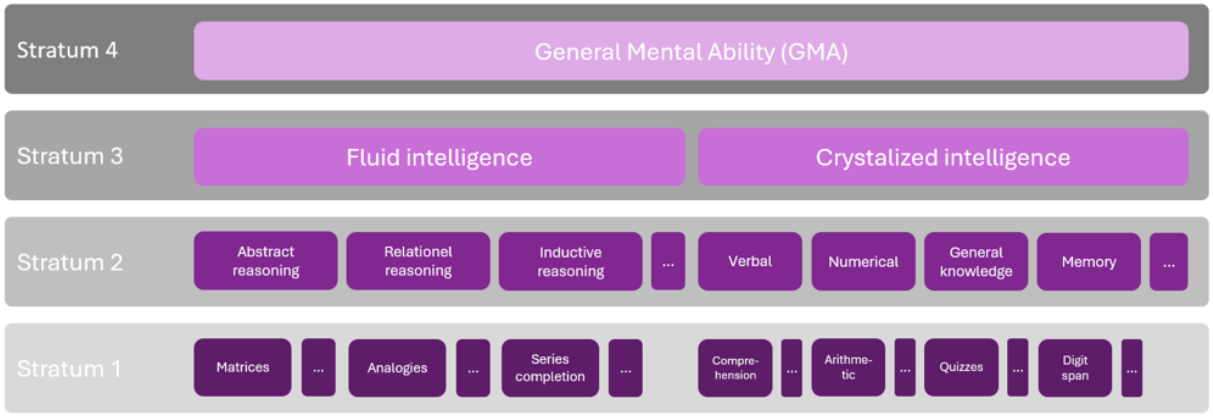
Content validity

Content validity concerns whether test items and scales are relevant and representative of the theoretical concept (domain) being measured.

Within classical intelligence theory, Charles Spearman's made the distinction between a general (g) factor and specific (s) factors. Later, Raymond Cattell's divided the concept of intelligence into fluid and crystallized intelligence. A hierarchical model of these dimensions is illustrated in Figure 5.1. These hierarchical layers (or strata) mirror the Cattell-Horn-Carroll Theory of Cognitive Abilities, which has received extensive empirical support throughout several years of research (Flanagan & Dixon, 2014).

Based on this model, Matrigma measures abilities in between stratum 2 and 3. Although figure-based items are usually considered a measure of abstract reasoning, the matrix format also involves pattern recognition and deductive reasoning. Therefore, it can be argued that Matrigma measures fluid intelligence more broadly, which in turn is related to General Mental Ability (GMA). Within research, assessments of this kind (e.g., Raven's Progressive Matrices) are also widely recognized as good indicators of both GMA and fluid intelligence in particular (Jensen, 1998).

Figure 5.1. Hierarchical model of cognitive abilities.



Construct validity

Construct validity concerns the agreement between test results and prior theoretical knowledge of the construct being measured.

Convergent validity (Logics)

One way to prove the validity of a test is by showing that it correlates with other tests measuring the same construct. This is also referred to as convergent validity. Thus, we invited 180 job candidates to complete both Adaptive Matrigma and Logics, which is another cognitive test offered by Assessio. Participants consented for their data to be used for research purposes.

As opposed to Adaptive Matrigma which only measures fluid intelligence, Logics also measures crystallized intelligence. To measure fluid intelligence, however, Logics relies on both figural and verbal items. It also accounts for answering speed as much as for accuracy. However, one should note that there seems to be a trade-off between the two. This means that the faster a candidate answers, the more mistakes they tend to make. Table 5.1 shows the correlations between Adaptive Matrigma and Logics.

Table 5.1. Correlations between Adaptive Matrigma and Logics.

Domain	Scale	Definition	Pearson*	Spearman
GMA	Cognitive Capacity	The ability to quickly and precisely solve different tasks.	.49	.51
	Learning Capability	The ability to learn new things and solve novel logical problems.	.49	.52
Decision Style	Speed	The speed at which decisions regarding solutions and tasks are made.	.24	.23
	Accuracy	The ability to conceive and derive correct solutions.	.50	.55
Logical Capacity	Perception	The ability to perceive and observe logically.	.45	.46
	Analysis	The ability to think and conclude logically.	.41	.46
	Complexity	Understanding and solving complex tasks.	.43	.46

Skills	Numerical	The sense of mathematics and calculations.	.37	.36
	Verbal	The sense of language and grammar and the ability to use the language correctly.	.28	.28

**Note*. N = 166 (14 outliers excluded).

Overall, this data shows moderate to high correlations between Adaptive Matrigma and Logics. Thus, we can conclude that both measure general mental ability. It is likely those correlations would have been even higher if the tests had the same format (e.g., if both were non-verbal). As expected, Adaptive Matrigma has stronger correlations with fluid intelligence components of Logics than with the crystallized intelligence components. In addition, the correlation with Speed is modest, which makes sense given that Adaptive Matrigma was not designed as a speed test but is adaptive.

Convergent & discriminant validity

In addition to the convergent validity reported above, we examined correlations between the GMA score and Learning Agility, Values, and Big Five personality traits (assessed with Match-V and MAP). Correlations are shown below in Table 5.2.

As expected, GMA correlated moderately with Change and Mental Agility, the values of Change, Curiosity, and Idealism and the personality trait Openness. The remaining correlations were small or near-zero supporting the discriminant validity of Adaptive Matrigma.

Table 5.2. Correlations between GMA, Learning Agility, Values, and Personality.

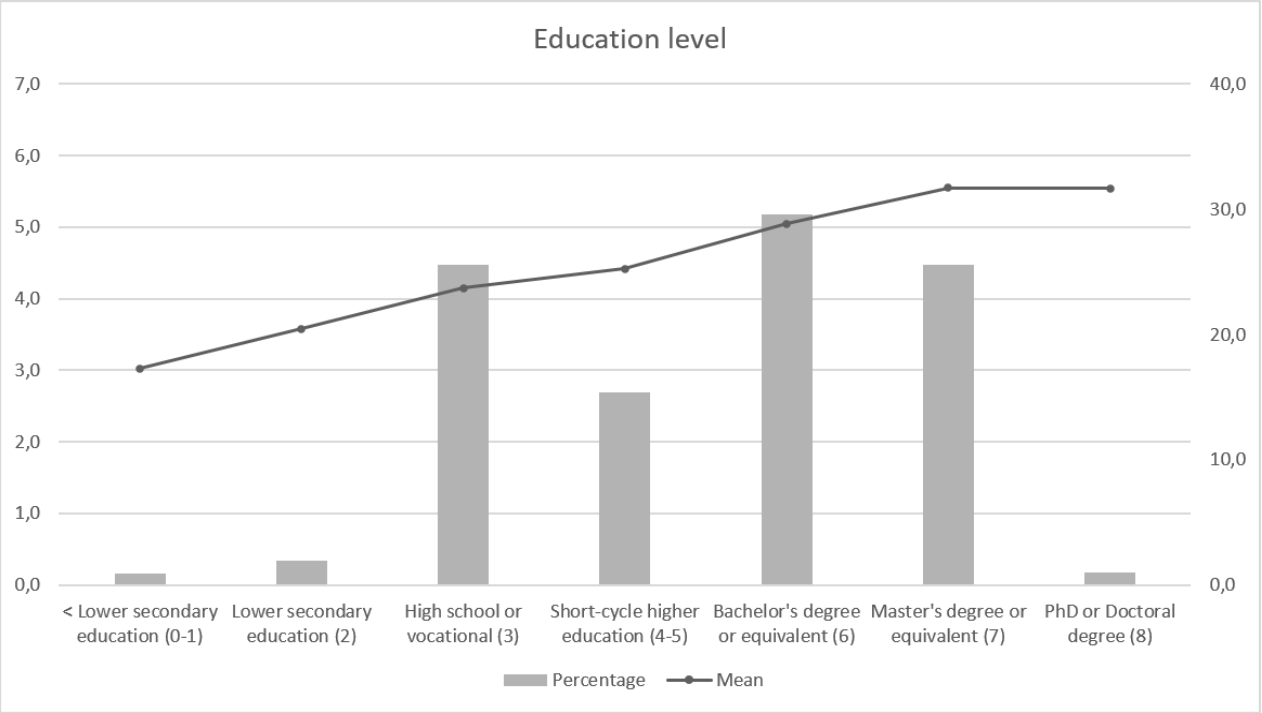
Construct	Dimension	GMA
Learning agility	Change agility	.35
	Mental agility	.36
	People agility	.11
	Results agility	.13
	Self-awareness	.25
	Total	.35
Values	Status	.01
	Achievement	.01
	Pleasure	.22
	Change	.30
	Curiosity	.35
	Idealism	.33
	Connection	.13
	Conformity	-.06
	Security	-.02
Personality	Extraversion	.21
	Agreeableness	.09
	Conscientiousness	.05
	Emotional Stability	.09
	Openness	.31

Education level

Besides these studies, the construct validity of Adaptive Matrigma is further supported by a relationship between the GMA scores and education level. Figure 5.2 shows the mean GMA score on the left axis and the percentage of cases on the right axis across different education levels. The numbers in parentheses reflect the level as defined by the International Standard Classification of Education (ISCED).

As expected, mean scores increase linearly with the level of education. When combined, the low to mid-level education levels (< Bahelor's degree) had a mean C score of 4.2, and the high levels education levels (Bahelor's degree or higher) had a mean C score of 5.3.

Figure 5.2. Mean GMA score across education levels defined by ISCED.



Criterion validity

This refers to how well the test predicts relevant performance criteria.

A small study of N = 21 associates, managers, partners, and directors was conducted in 2025 studying the relationship between GMA and five different outcomes (criteria):

- **Task Performance (TP)**: Completing tasks and meeting formal requirements of the job.
- **Organizational Citizenship Behavior (OCB)**: Behaviors outside of formal role requirements that help or benefit team members or the organization more broadly.
- **Adaptive Work Behavior (AWB)**: Being innovative and adapting to new circumstances and unforeseen events.
- **Total performance**: Average of TP, OCB, and AWB.
- **Net Promotor Score (NPS)**: The likelihood of the person hiring a similar profile for the job in question.

As preliminary analyses showed different prediction patterns for the group of managers and associates vs. partners and directors, the analysis was split by function. Results are reported only for the first group.

Importantly, a notable range restriction was observed on GMA scores in the sample. In recognizing this is a concurrent study, this range restriction is likely to be only indirect rather than direct (hence warranting more conservative estimates). To correct for this and criterion unreliability, we applied the following parameters and assumptions:

- Criterion reliability was estimated at 0.80 as inter-rater agreement was quite high within the respective group of raters (manager's managers, immediate managers, peers).
- The ratio of restricted to unrestricted SDs was set at U = 0.58.
- A modest relationship was assumed between performance (y) and the initial selection variable (z) with a correlation of r(y,z) = 0.20.
- A sensitivity analysis was made to estimate the correlation (or overlap) between GMA (x) and the initial selection variable (z):
    - Low: r(x,z) = .20 (like unstructured interviews)
    - Moderate: r(x,z) = .30 (like work samples or assessment centers)
    - Strong: r(x,z) = .45 (like Grade Point Average, GPA)

Table 5.2 shows the observed correlations and validity estimates for each of the performance dimensions. Surprisingly, the strongest correlations with GMA were observed for OCB, whereas correlations with TP were only modest. However, this is likely due to a job effect with the job being very relational in nature. For total performance and NPS, the observed correlation was .25, and the validity estimates ranged from .35 to .40.

Table 5.2. Correlations and validity estimates for GMA across performance dimensions.

GMA	Overlap	TP	OCB	AWB	Total	NPS
Observed correlation	-	.04	.33	.10	.25	.25
Validity estimate	Low	.12	.45	.19	.35	.35
	Moderate	.15	.46	.22	.37	.37
	Strong	.19	.46	.25	.40	.40

6. Reliability

Reliability is defined as the consistency with which an instrument measures a construct.

As Adaptive Matrigma was developed on the basis of IRT, traditional reliability coefficients (such as Cronbach's alpha) cannot be used to summarize the overall reliability of the test. Instead, we compared scores between Classic and Adaptive Matrigma. Thus, the correlation between the two assessment is used as an estimate of parallel/alternate forms reliability.

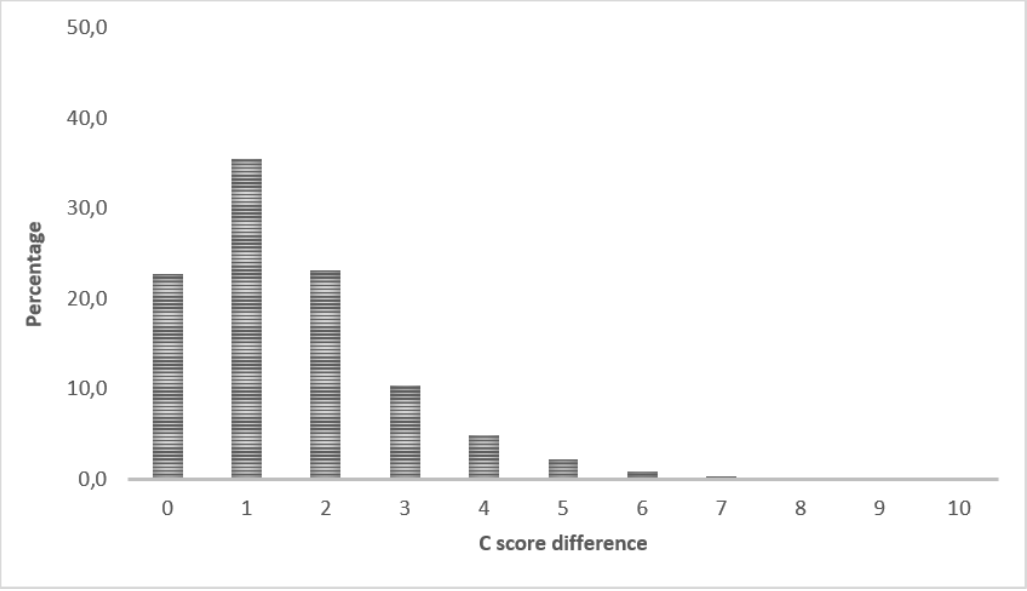
Table 6.1 shows the Pearson and Spearman correlations between scores from Classic and Adaptive Matrigma in a sample of N = 10,182 candidates who completed both assessments mostly (93%) in a high-stake (recruitment) setting. Both correlations are reported for the full sample and a reduced sample excluding a total of 367 outliers (3.6 % of the sample). For the full sample, the correlation was .57, which increased to .68 representing a strong correlation and almost adequate reliability.

Table 6.1. Correlations between Adaptive and Classic Matrigma.

GMA (Adaptive/Classic)	N	Pearson	Spearman
Total sample	10 182	.57	.54
Excluding outliers	9 815	.68	.58

Besides correlations, we also examined score differences at the C score level between GMA measured by Adaptive and Classic Matrigma depicted in Figure 6.1. For absolute values, 58% had a C score difference of 0-1 C score, 33% had a difference of 2-3 C scores, while the remaining 8.5% had a C score difference of 4 points or higher.

Figure 6.1. C score differences between Adaptive and Classic Matrigma.



7. Standardization

As mentioned in an earlier chapter, Adaptive Matrigma was developed based on Classic Matrigma. Throughout the years, Classic Matrigma was updated multiple times. Adaptive Matrigma is based on the 2016 update, which is also the most recent one. Thus, the difficulty, discrimination, and guessing rate of Adaptive Matrigma items were calculated based on the 2016 norm. This included a sample of 4,606 participants.

Table 7.1 shows the composition in terms of gender, age, educational level, and completion language.

Table 7.1. Demographic characteristics of the norm group for Classic Matrigma.

Characteristic	n	%	M	SD
Age	4606	-	37	11
Gender				
Women	2050	45		
Men	2556	55		
Educational level				
Mandatory education	70	2		
Highschool	1666	36		
Bachelor's degree	1511	33		
Master's degree	1267	28		
PhD or similar	92	2		
Swedish	3431	74		
English	621	13		
Norwegian	434	9		
Finnish	81	1.8		
Danish	25	0.5		
German	11	0.2		
French	3	0.1		

Group Differences and Adverse Impact

When using an assessment to make important decisions with a great impact on individuals (such as selection, promotion, and hiring decisions), a key requirement is to ensure fairness and mitigate Adverse Impact (AI), defined as "a substantially different rate of selection in hiring, promotion, or other employment decisions which works to the disadvantage of members of a race, sex or ethnic group" (Uniform Guidelines on Employee Selection Procedures, Equal Employment Opportunity Commission, 1978). The "Four-Fifths rule" can be used to determine whether an assessment has AI. Usually, a selection rate for any demographic group less than four-fifths (or 80 percent) of the selection rate for the group with the highest rate (majority group) is considered evidence of AI. The level of AI depends both on the magnitude of group differences (e.g., between men and women) and the selection ratio, i.e., the number of people hired compared to the total number of applicants. Therefore, we found it important to investigate whether Adaptive Matrigma has any adverse impact.

Simulations of expected AI were conducted at different selection rates for gender (men/women) and age (below/above 40). Please note, however, that these calculations are based on the assumptions that: 1) candidates are selected based on a single score only, 2) the

assessment is used as the sole basis for selection, and 3) a fixed selection rate is applied (i.e., hiring everyone scoring above a predefined cut-off). In practice, Assessio recommends basing recruitment decisions on a combination of assessments, scales, and other information relevant to the job in question (i.e., KSAOs) to consider both job, team, and organization fit. Generally, combining methods decreases adverse impact but not if all methods negatively impact the same group. Hence, it is important to be mindful of potential adverse impact at a scale level, even with the recommended application of assessments.

Tables 7.2 and 7.3 list the standardized mean difference (Cohen's d) between groups alongside the simulated AI ratio (selection rate of the least represented group compared to the most represented group) for gender (men/women), and age (below/above 40), respectively. The calculations are based on three fixed selection ratios (SR): Strict (C score 7-10), Moderate (C score 6-10) and Lenient (C score 5-10) equivalent to the top 23, 40, and 60%, respectively. We aimed for an AI ratio above 0.80 for a lenient selection ratio as suggested by the Four-Fifths rule.

For gender, no adverse impact is expected when using Adaptive Matrigma. AI ratios far exceeded the 0.80 threshold, regardless of the selection ratio used. For age, careful consideration should be paid to a potential adverse impact favoring younger individuals when applying a strict selection ratio. If a strict selection ratio must be applied, then we recommend combining Adaptive Matrigma with other assessments (e.g., personality or values) or criteria to balance out the level of Adverse Impact. This would prevent any indirect discrimination of older applicants.

Table 7.2. Adverse Impact simulations for gender (men/women) at different selection ratios (SR).

Scale	d	Strict SR	Moderate SR	Lenient SR
GMA	0.003	0.91	1.00	0.98

Table 7.3. Adverse Impact simulations for age (below/above 40) at different selection ratios (SR).

Scale	d	Strict SR	Moderate SR	Lenient SR
GMA	0.23	0.74	0.83	0.91

8. Translations and Adaptations

Adaptive Matrigma is a non-verbal test. Candidates don't need to make use of language to answer the questions. However, sufficient proficiency in the test language in necessary to understand the instructions and feedback report. We make efforts to ensure that the test is available in the most requested languages. To see which languages are available at the moment, please reach out to your contact person in Assessio.

References

DeMars, C. ((2019). Item response theory. Understanding statistics measurement. Oxford, New York: Oxford University Press.

Furr, R.M. (2011). Scale Construction and Psychometrics for Social and Personality Psychology. SAGE Publications. ISBN 9780857024046.

Jensen, A. R. (1998). The g Factor. The science of Mental Ability. Westport, CT: Praeger Publisher.

Kehoe, J.F., & Sackett, P.R. (2010). Validity consideration in the design and implementation of selection systems. In Farr, J.L., & Tippins, N.T (Eds.). Handbook of employee selection (pp. 56-92). New York, NY: Routledge.

Mabon, H. (2014). Arbetspsykologisk testning. Assessio International: Stockholm, Sverige.

Mabon, H., & Sjöberg, A. (2016). Matrigma. Technical Manual. Stockholm: Assessio International.

Raven, J. C. (1998). Raven's progressive matrices and vocabulary scales. Oxford Psychologists Press.

Ryan, J. J., Sattler, J. M., & Lopez, S. J. (2000). Age effects on Wechsler adult intelligence scale-III subtests. Archives of Clinical Neuropsychology, 15(4), 311-317.

Schmidt, F. L., Oh, I. S., & Shaffer, J. A. (2016). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 100 Years of Research Findings. Fox School of Business Research Paper.

Suggested readings on Item Response Theory:

Baker, F. (2001). The Basics of Item Response Theory. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD.

Baker, F. B., & Kim, S-H (2004). Item Response Theory: Parameter Estimation Techniques (2nd ed.). Marcel Dekker. ISBN 978-0-8247-5825-7.

de Boeck, P., & Wilson, M. (2004). Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach. New York, NY: Springer. ISBN 978-0-387-40275-8.

Embretson, S. E., & Reise, S. P. (2000). Item Response Theory for Psychologists. Psychology Press. ISBN 978-0-8058-2819-1.

Fox, J-P. (2010). Bayesian Item Response Modeling: Theory and Applications. New York, NY: Springer. ISBN 978-1-4419-0741-7.

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Mahwah, NJ: Erlbaum.

van der Linden, W. J., & Hambleton, R. K., eds. (1996). Handbook of Modern Item Response Theory. New York, NY: Springer. ISBN 978-0-387-94661-0.