Face and Voice Recognition Lab
School of Human Sciences
Institute of Lifecourse Development
University of Greenwich
London
www.superrecognisers.com
super-recognisers@greenwich.ac.uk
www.twitter.com/GREcognisers

Version 5: 30 April 2024

University of Greenwich Face and Voice Recognition Lab

Volunteer Research Participant Pool

Blog 2: How do your scores on our tests compare with those by volunteer pool members and members of the population?

Table of contents

1. Identity processing test scores: -
- Test 1: Could You be a Super-Recogniser Test (CYBSRT)
- Test 2: Cambridge (Male) Face Memory Test: Extended (CFMT+)
- Test 3 Glasgow Face Matching Test (GFMT)
- Test 4 Short-Term Face Memory Test 30-60 (STFMT3060)
- Test 5 Kent Face Matching Test (KFMT)
- Test 6 Cambridge Cars Memory Test (CCarMT)
- Test 7 Glasgow Face Matching Test: Version 2 (High 40) (GFMT2HI40)
- Test 8 Cambridge (Female) Face Memory Test (CFemFMT)
- Test 9 Cambridge Face Memory Test for Young People (CFMTYP)
- Test 10 University of New South Wales Test (UNSWT)[1]
- Test 11 Bangor Voice Memory Test (BVMT)
- Test 12 Glasgow Voice Matching Test (GVMT)
- Test 13 Greenwich Voice Memory Test 88: Version 1 (GreVMT88)[2]
- Test 14 Greenwich Voice Memory Test 104 (GreVMT104)
- Test 15 Jena Voice Learning and Memory Test (JVLMT)[3]

1 Test available on the University of New South Wales website only. We would be grateful that if you take this test, you please enter you score into the UNSW response form on the Greenwich Test and Research Distribution Platform.
2 Test not currently available on the Greenwich Test and Research Distribution Platform. It is due to be reactivated in the next few weeks.
3 Test not currently available on the Greenwich Test and Research Distribution Platform. It is due to be reactivated later in 2024.

2. Defining face recognition ability for research.

Appendix B: Tests included in this blog.
List of References.

Identity processing tests: Results

The distributions of the test scores for each of the main tests taken by our volunteer pool are shown in the following figures. You should be able to compare your own scores with these to get a good idea of your ability. You will find the minimum thresholds we use when deciding whether someone is a super-recogniser or not for research purposes. Depending on research design and the number of tests used to estimate ability, we may use strict super-recogniser criteria (Test scores approximately 2 standard deviations (SD) above the typical mean) or liberal super-recogniser criteria (Test scores approximately 1.5 SD above the typical mean). Table 2 provides an example of how the data is reported for each test.

Table 3: Guide to test data reports

A histogram describing the spread of scores is provided for each test. We mostly report 1st and 2nd attempt scores only. In the same section we report z-scores for each test based on mean performances on that test by a population-representative sample of participants.

μ (S)	Estimated population-representative mean score (z = 0)
z = - 1.00	1.0 SD below mean score (lower typical-ability control threshold)
z = +1.00	1.0 SD above mean score (upper typical-ability control threshold)
z = +1.50	1.5 SD above mean score (liberal super-recogniser threshold)
z = +2.00	2.0 SD above mean score (strict super-recogniser threshold)
z = +x.xx	?.? SD above mean score (maximum test score)

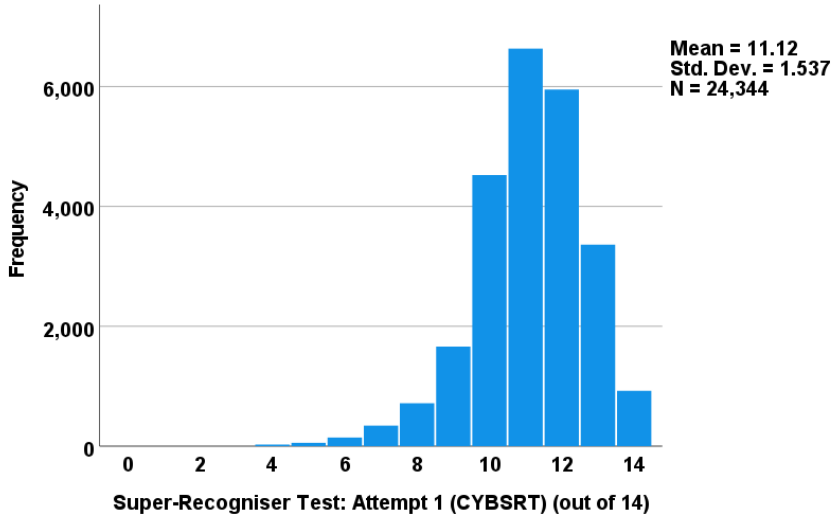
Super-recogniser thresholds: A score on a test that is at least 2 Standard Deviations (SD) above the population-representative mean has been the standard for super-recogniser group membership, since the first research on this skill was conducted by Russell et al. (2009). It matches the typical standard of 2 SDs below the mean for developmental prosopagnosia, which is sometimes called ‘face blindness’. About 2.5% of the population would be expected to achieve a score in this range on a single test. Substantially more (6.5%) will achieve our 1.5 SD above the mean liberal standard for super-recognition.

Typical ability-level control thresholds: A far higher proportion of the population (68%) is expected to achieve a score that falls within 1 SD of the mean. This has often been our typical-ability control standard in research.

Why do we employ strict and liberal criteria? Although 2.5% of the population would be expected to achieve a score 2 SD above the mean on a single test, if the > 2 SD standard were to be applied to two tests, fewer participants would achieve criteria. Even fewer will achieve the criteria if three tests are used, and so on. Therefore, if multiple tests are used to define eligibility, it may be necessary to apply liberal criteria, to ensure numbers of participants recruited will generate sufficient statistical power.

Test 1: Could You be a Super-Recogniser Test (CYBSRT)

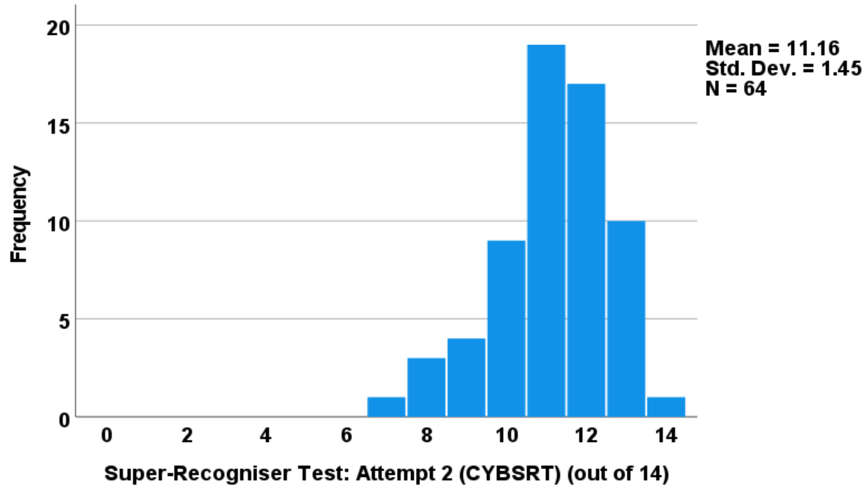
Figure 2.1: Could you be a Super-Recogniser Test: First attempts.



The normative data for this test were first described in Davis (2019)

Could you be a Super-recogniser Test	Score	SD
μ	9.64	(1.90) (z = 0)
z = -1.00	7.74	Lower limit for typical-ability controls
z = +1.00	11.54	Upper limit for typical-ability controls
z = +1.50	12.49	Liberal Super-Recogniser Threshold
z = +2.00	13.44	Strict Super-Recogniser Threshold
z = +2.29	14.00	Maximum score

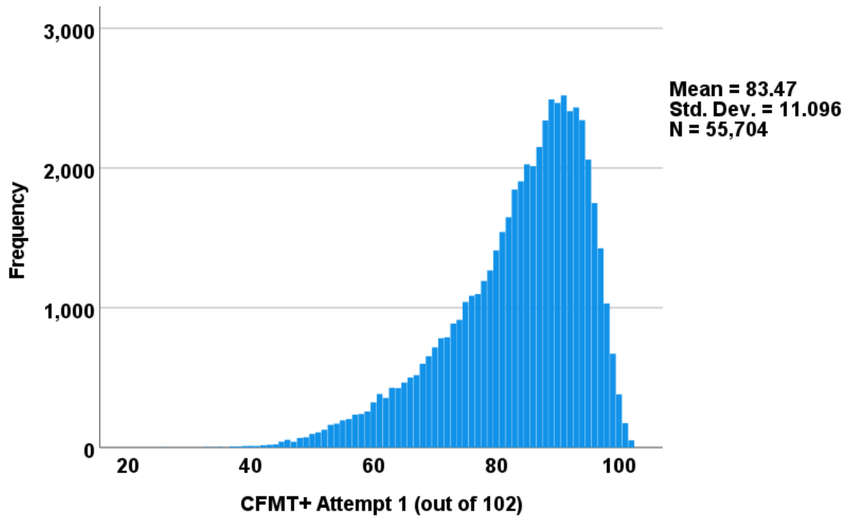
Figure 2.2: Could you be a Super-Recogniser Test: Second attempts.



Data from just over 24,000 participants are reported. Many more have taken this test. However, no identity information is collected when participants first take the tests and to retain data, participants must manually enter their score into the system. About 75% do this. However, when we first created the volunteer pool, this manual entry system was not set up.

Test 2: Cambridge (Male) Face Memory Test: Extended (CFMT+)

Figure 3.1: Cambridge (Male) Face Memory Test: Extended (CFMT+): First attempts.

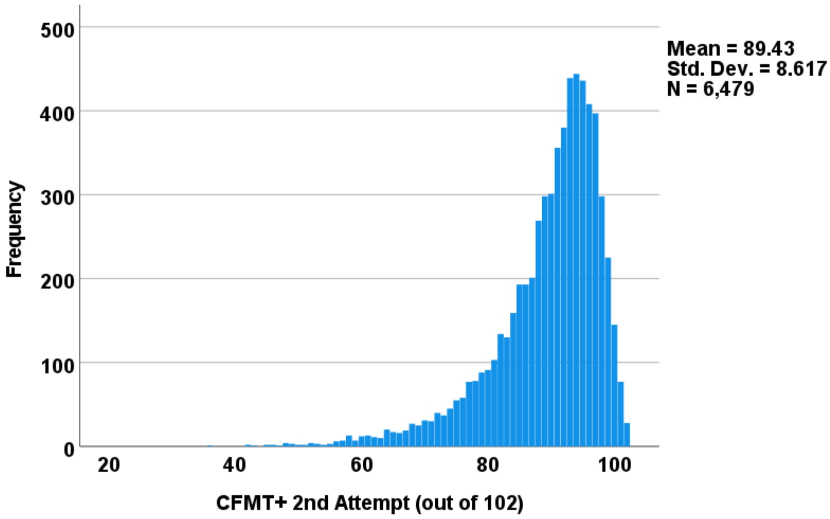


CFMT+ population norms (*n* = 254) were extracted from Bobak et al. (2016).

	Score	SD
μ (S)	70.70	(12.30) (z = 0)
z = -1.00	58.40	Lower limit for typical-ability controls
z = +1.00	83.00	Upper limit for typical-ability controls
z = +1.50	89.15	Liberal Super-Recogniser Threshold

z = +2.00	95.30	Strict Super-Recogniser Threshold
z = +2.54	102.00	Maximum score

Figure 3.2: Cambridge (Male) Face Memory Test: Extended (CFMT+): Second attempts.



Strict standards for super-recognition: To understand how we use these data, in most of our research to be classified as a super-recogniser a participant must achieve a score of 95 on this test (2 SD is actually equivalent to a score of 95.3, but we round down incremental values to whole numbers. We call this our ‘strict’ or conservative super-recognition criterion for this test. However, different published papers by other researchers have reported lower super-recognition thresholds based on scores 2 SD above the mean (i.e., 88-90 out of 102). These outcomes are due to lower sample mean scores or reduced sample variance.

Liberal standards for super-recognition: We also report liberal super-recognition criteria of 1.5 SD above the mean on each test. This lower threshold may be used if multiple tests are employed, as increasing the number of tests, inevitably reduces the proportion of participants who achieve the 2.0 SD above the mean standard. On the CFMT+, the value of 89.3 is typically rounded down to 89.0.

Typical-ability control standards: Researchers have sometimes assigned all participants who do not achieve super-recognition standards to control groups, on the assumption that they form a representative sample of non-super-recognisers. This would be inappropriate in most research in which random samples of the Greenwich Volunteer Pool are recruited, due the tendency for members to possess superior face recognition ability. Even if participants do not achieve super-recogniser standards, there is an over-representation of participants who score fractionally below that standard (see Table 1). Evidence for this is that the mean score of participants in Bobak et al.’s (2016) population-representative sample was 70.7 (SD = 12.3), while the volunteer pool’s equivalent is 83.47 (SD = 11.10) on their first attempt, and 89.43 (SD = 8.62) on their second.

These values are approximately 1 SD and 1.5 SD above Bobak et al.’s (2016) mean scores.

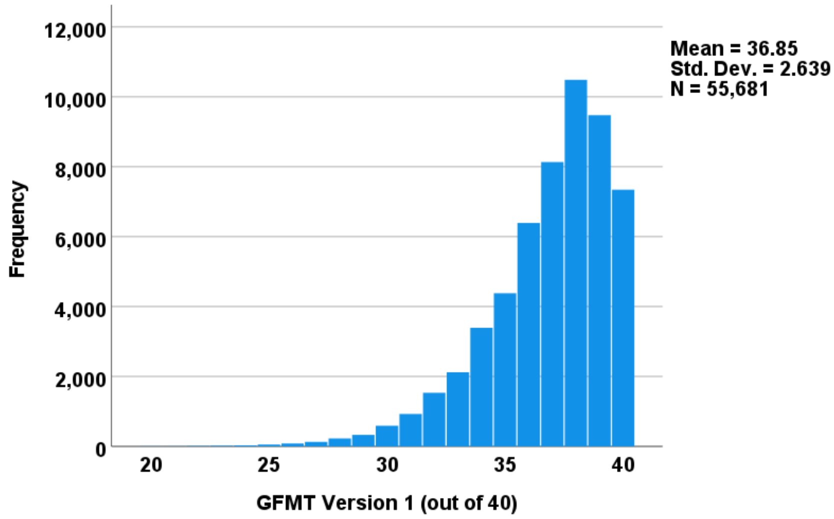
Therefore, when we have allocated participants to control groups, we have employed one of two strategies.

The first strategy involves selecting participants who generate scores within 1 SD of the population-representative sample mean, on the basis that they can be labelled as possessing typical-ability skills. In the case of the CFMT+, those achieving scores from 58-83 are assigned to control groups. However, the bias for the Face and Voice Recognition Lab to recruit participants with superior skills extends throughout the entire volunteer pool. Control groups allocated using the first strategy, inevitably present with mean scores that are higher than the representative sample mean.

The second strategy is therefore to extend the lower threshold from 1 standard deviation below the mean to 1.5 standard deviations below the mean, as the lower threshold for control group membership.

Test 3: Glasgow Face Matching Test (GFMT)

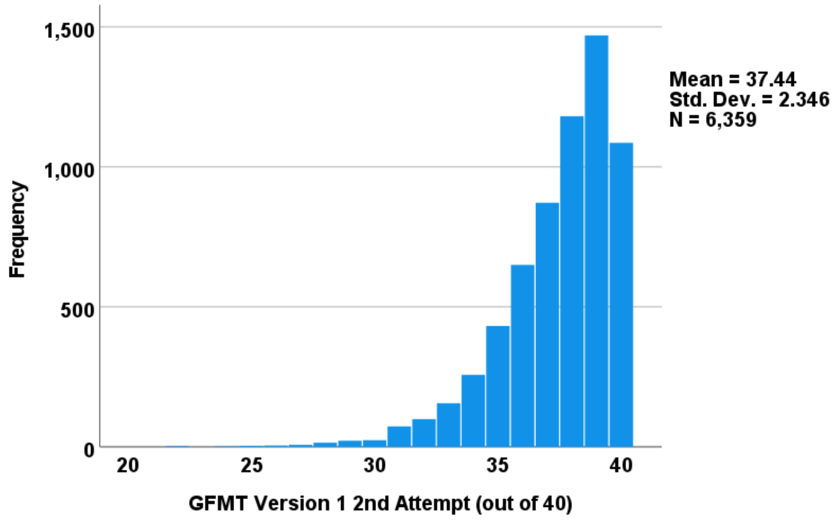
Figure 4.1: Glasgow Face Matching Test (GFMT): First attempts.



GFMT population norms (*n* = 194) were first reported in Burton et al. (2010).

	Score	SD
μ (S)	32.50	(3.90) (z = 0)
z = - 1.00	28.60	Lower limit for typical-ability controls
z = +1.00	36.40	Upper limit for typical-ability controls
z = +1.50	38.35	Liberal Super-Recogniser Threshold
z = +2.00	40.30	Strict Super-Recogniser Threshold
z = +1.92	40.00	Maximum score

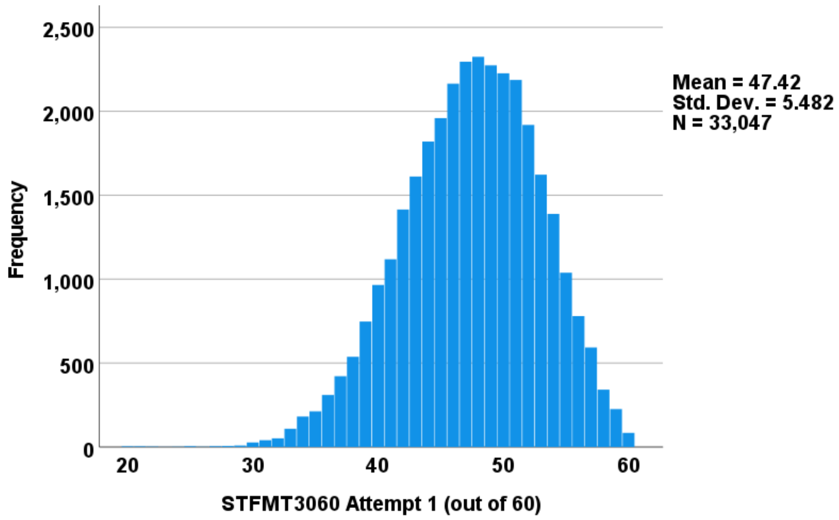
Figure 4.2: Glasgow Face Matching Test (GFMT): Second attempts.



A score 2 SD above the mean on this test is impossible to achieve. However, as with the other tests, we round the value of 40.3 down to the nearest whole number (40 out of 40) as indicative of strict super-recognition ability. Because of this, in recent years, we have placed more weight on the results of the Kent Face Matching Test, and the Glasgow Face Matching Test Version 2 when assigning participants to super-recogniser groups in research.

Test 4: Short-Term Face Memory Test 30-60 (STFMT3060)

Figure 5.1: Short-Term Face Memory Test 30-60 (STFMT3060): First attempts.

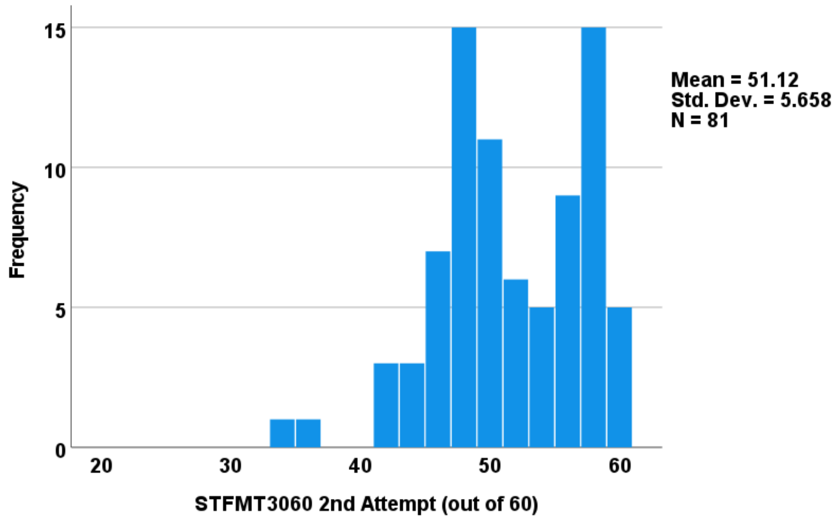


STFMT3060 population norms were calculated from a sample of the Volunteer Pool who had achieved scores within 1 standard deviation of the mean on at least three other tests.

	Score	SD
μ (S)	45.17	(4.31) (z = 0)
z = - 1.00	40.86	Lower limit for typical-ability controls
z = +1.00	49.48	Upper limit for typical-ability controls

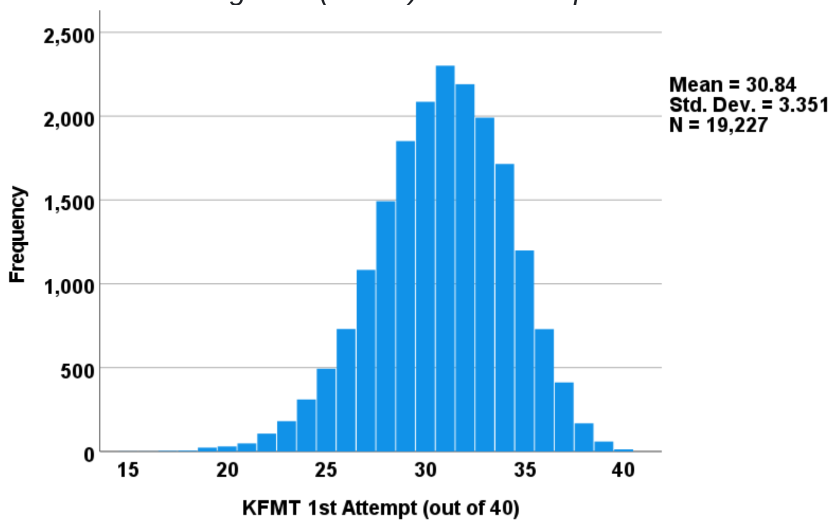
z = +1.50	51.64	Liberal Super-Recogniser Threshold
z = +2.00	53.79	Strict Super-Recogniser Threshold
z = +3.44	60.00	Maximum score

Figure 5.2: Short-Term Face Memory Test 30-60 (STFMT3060): Second attempts.



Test 5: Kent Face Matching Test (KFMT)

Figure 6.1: Kent Face Matching Test (KFMT): First attempts.

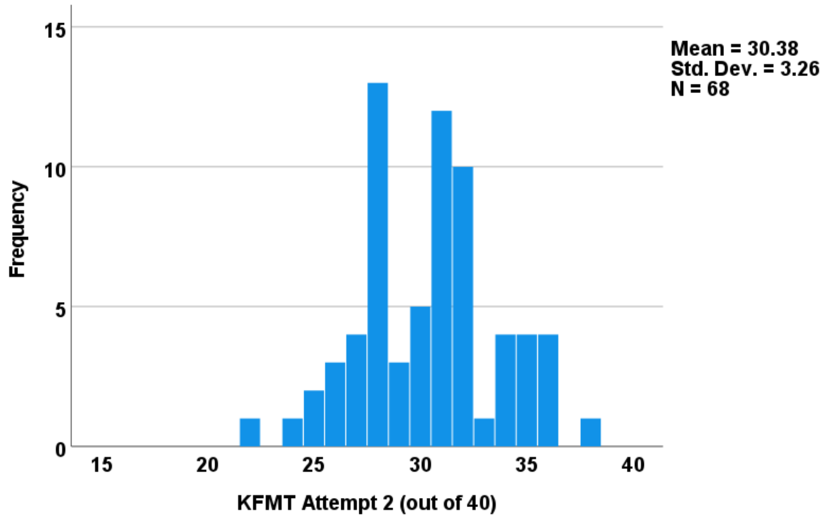


KFMT population norms were compiled from three articles (Fysh, 2018; Fysh & Bindemann, 2018; Gentry & Bindemann, 2019).

	Score	SD
μ (S)	26.50	(3.45) (z = 0)

z = - 1.00	23.05	Lower limit for typical-ability controls
z = +1.00	29.95	Upper limit for typical-ability controls
z = +1.50	31.67	Liberal Super-Recogniser Threshold
z = +2.00	33.40	Strict Super-Recogniser Threshold
z = +3.91	40.00	Maximum score

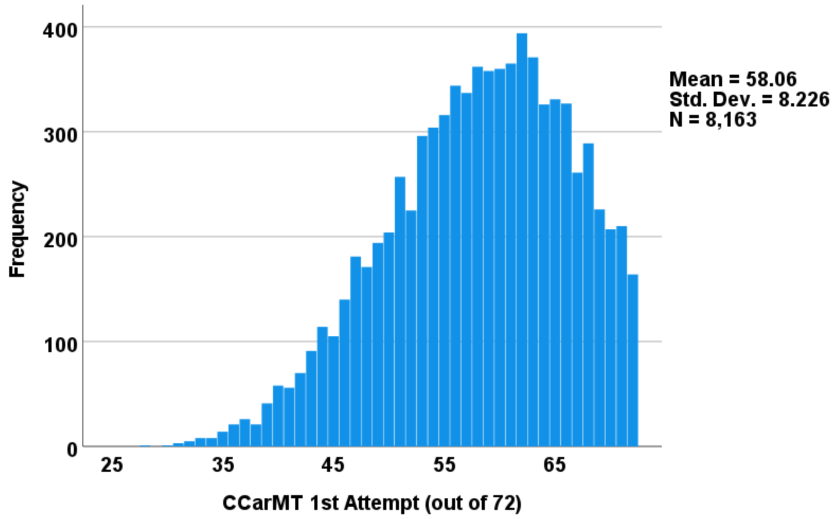
Figure 6.2: Kent Face Matching Test (KFMT): Second attempts.



Test 6: Cambridge Cars Memory Test (CCarMT)

CCarMT "population norms" (*n* = 153) were reported in Dennett et al. (2011).

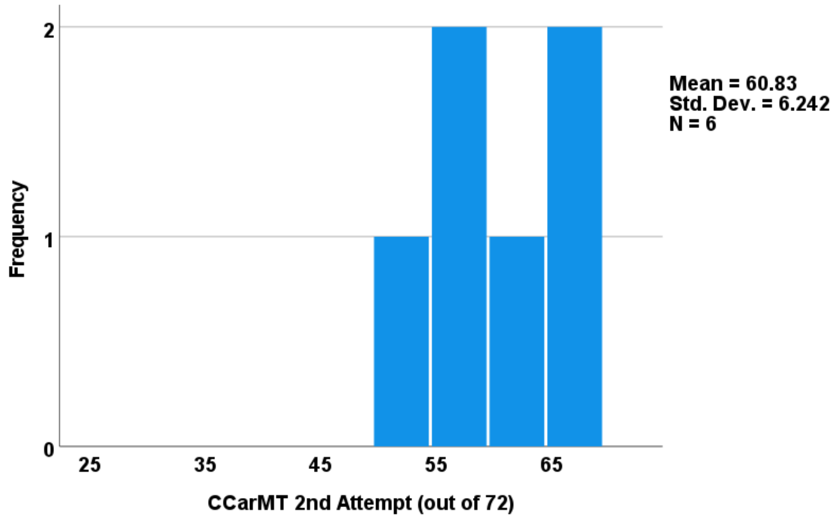
Figure 7.1: Cambridge Cars Memory Test (CCarMT): First attempts.



CCarMT population norms were reported in Dennett et al. (2011)

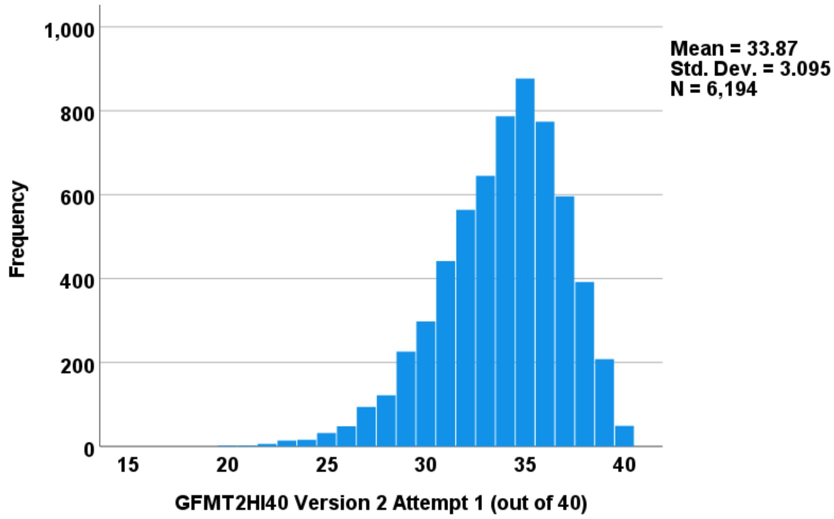
	Score	SD
μ (S)	53.18	(8.33) (z = 0)
z = - 1.00	44.85	Lower limit for typical-ability controls
z = +1.00	61.51	Upper limit for typical-ability controls
z = +1.50	65.68	Liberal Super-Recogniser Threshold
z = +2.00	69.84	Strict Super-Recogniser Threshold
z = +2.26	72.00	Maximum score

Figure 7.1: Cambridge Cars Memory Test (CCarMT): Second attempts.



Test 7: Glasgow Face Matching Test: Version 2 (High 40) (GFMT2HI40)

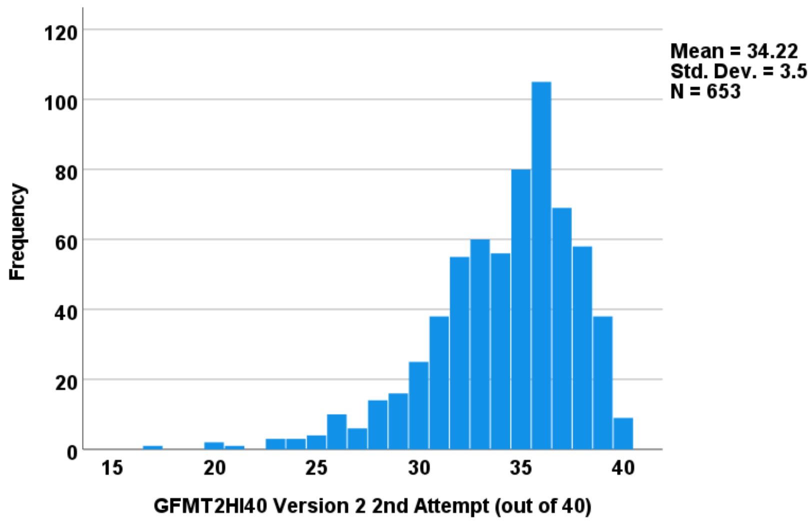
Figure 8.1: Glasgow Face Matching Test: Version 2 (High 40) (GFMT2HI40): First attempts.



GFMT2HI40 population norms were derived from the values reported in White et al. (2022)

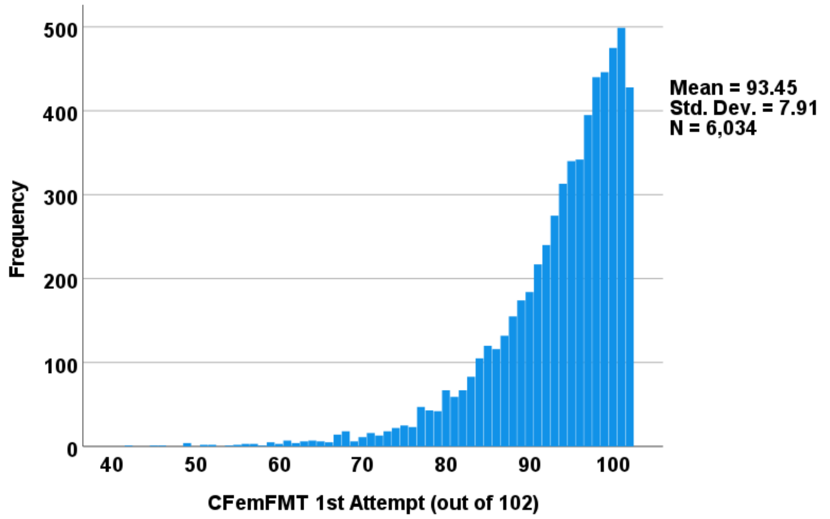
	Score	SD
μ (S)	30.65	(3.10) (z = 0)
z = -1.00	27.55	Lower limit for typical-ability controls
z = +1.00	33.75	Upper limit for typical-ability controls
z = +1.50	35.30	Liberal Super-Recogniser Threshold
z = +2.00	36.85	Strict Super-Recogniser Threshold
z = +3.02	40.00	Maximum score

Figure 8.2: Glasgow Face Matching Test: Version 2 (High 40) (GFMT2HI40): Second attempts.



Test 8: Cambridge (Female) Face Memory Test (CFemFMT)

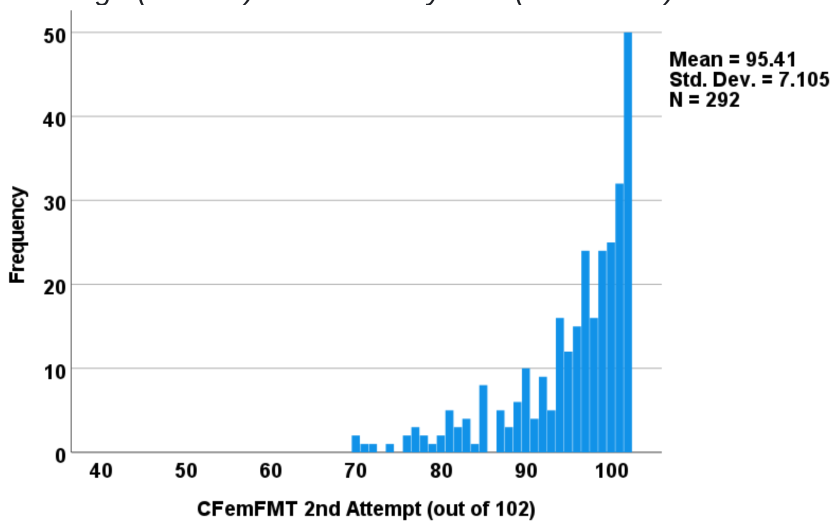
Figure 9.1: Cambridge (Female) Face Memory Test (CFemFMT): First attempts.



CFemFMT population norms were based on the data from Arrington et al. (2022)

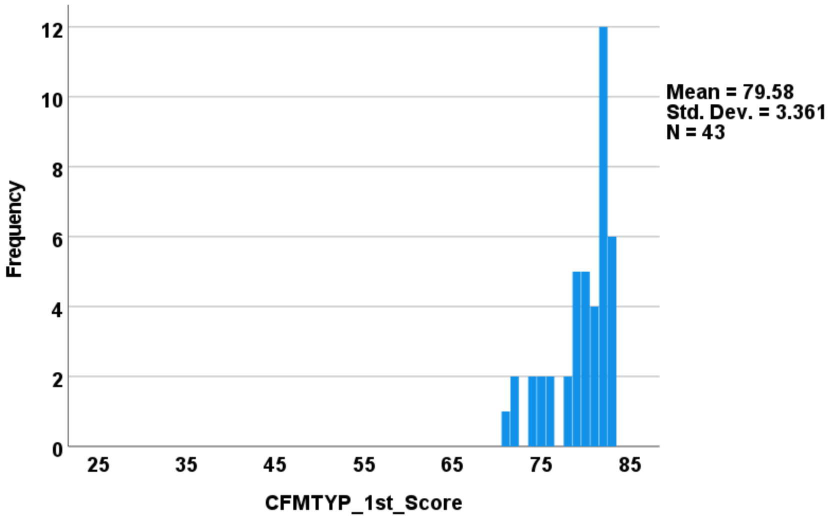
	Score	SD
μ (S)	81.63	(11.49) (z = 0)
z = -1.00	70.14	Lower limit for typical-ability controls
z = +1.00	93.12	Upper limit for typical-ability controls
z = +1.50	98.87	Liberal Super-Recogniser Threshold
z = +2.00	104.61	Strict Super-Recogniser Threshold
z = +1.77	102.00	Maximum score

Figure 9.2: Cambridge (Female) Face Memory Test (CFemFMT): Second attempts.



Test 9: Cambridge Face Memory Test for Young People (CFMTYP)

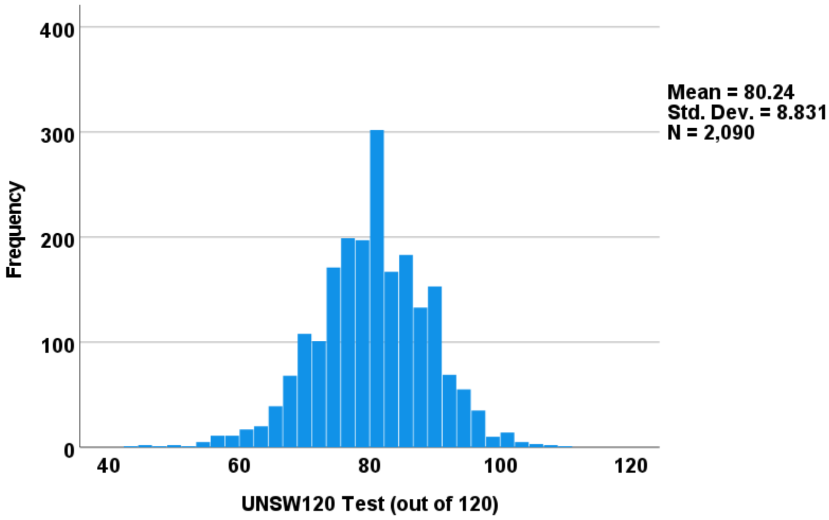
Figure 10.1: Cambridge Face Memory Test for Young People (CFMTYP): First attempts.



Note: This test was specially designed for the Greenwich Face and Voice Recognition Lab and until recently, has almost entirely been taken by children under 18-years. Until more adult participants have completed the test, it is not possible to generate normative data.

Test 10: University of New South Wales Test (UNSWT)

Figure 11.1: University of New South Wales Test (UNSWT): First attempts.



UNSW120 Test population norms were compiled from Dunn et al. (2021)

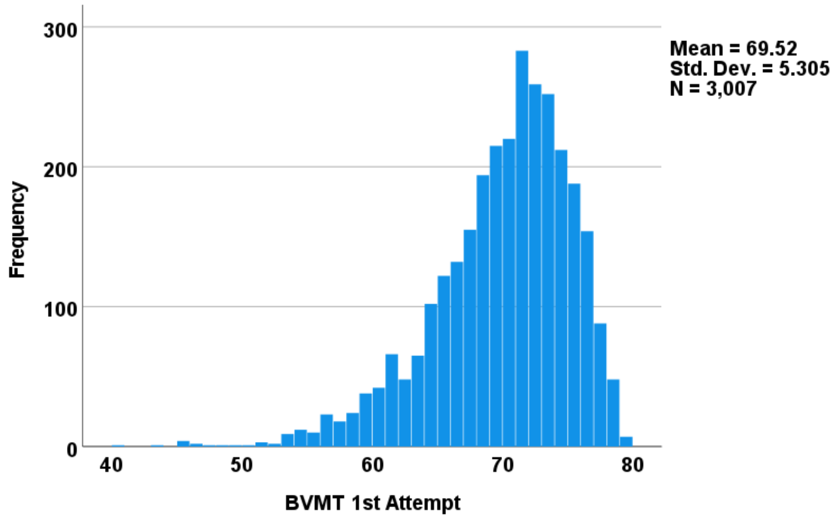
	Score	SD
μ (S)	70.68	(6.96) (z = 0)
z = -1.00	63.72	Lower limit for typical-ability controls
z = +1.00	77.64	Upper limit for typical-ability controls
z = +1.50	81.12	Liberal Super-Recogniser Threshold
z = +2.00	84.60	Strict Super-Recogniser Threshold
z = +7.09	120.00	Maximum score

The UNSW test was available for about a year on our website but was withdrawn due to interconnectivity failures with some browsers (the test itself was not the problem here). Prior to this, members of the Greenwich Volunteer Pool were invited to take the test loaded on the UNSW platform (see Dunn et al., 2021), and new volunteers have often informed us of their scores on this test. We have always recorded this information. This blog delivers the first public glimpse of these data. A dedicated link on the Test and Research Distribution Platform allows anyone to enter their score on this test (you may also upload a screenshot displaying the score as weightier proof of achievement). We aim to correlate scores on the UNSW test with those on other tests as part of the validation process.

It should be noted that maximum scores on the UNSW test are linked with the highest possible z-scores, although at the time of writing, to the best of our knowledge, no one has achieved a maximum score. The team who developed the test regularly make announcements about the highest scores.

Test 11: Bangor Voice Memory Test (BVMT)

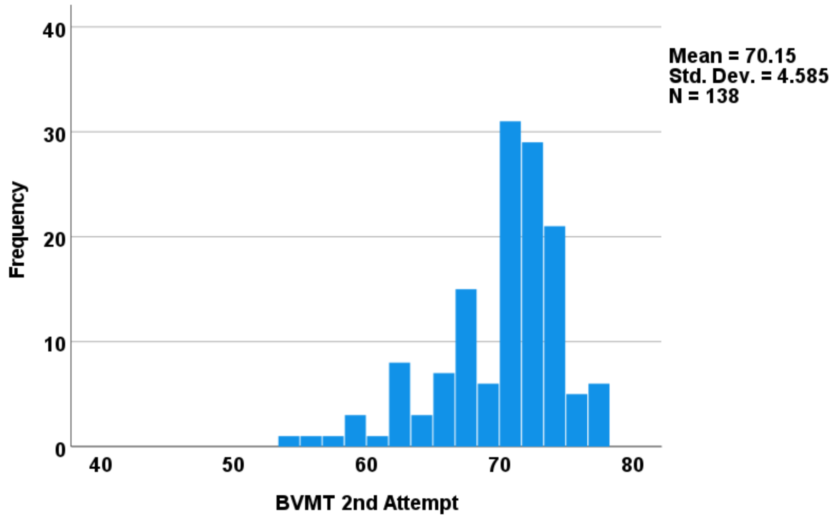
Figure 12.1: Bangor Voice Memory Test (BVMT): First attempts.



BVMT “population norms” were compiled from the data reported in Mühl et al. (2018)

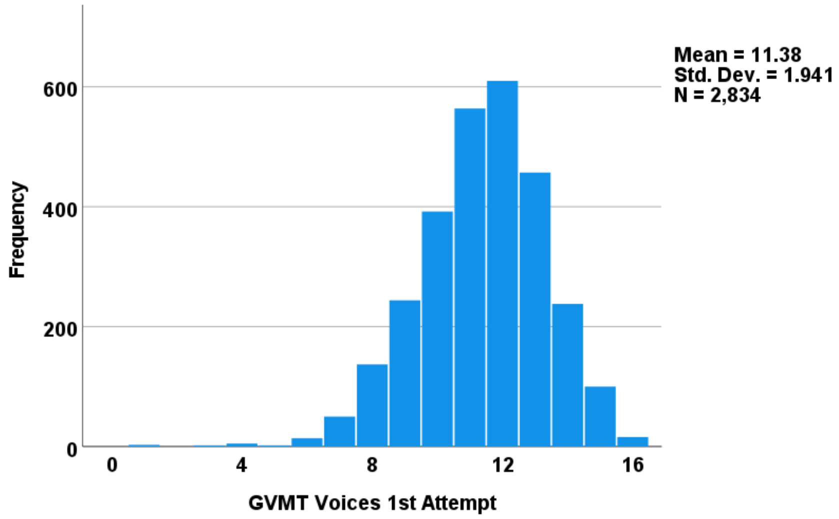
	Score	SD
μ (S)	67.66	(5.76) (z = 0)
z = -1.00	61.90	Lower limit for typical-ability controls
z = +1.00	73.42	Upper limit for typical-ability controls
z = +1.50	76.30	Liberal Super-Recogniser Threshold
z = +2.00	79.18	Strict Super-Recogniser Threshold
z = +2.14	80.00	Maximum score

Figure 12.2: Bangor Voice Memory Test (BVMT): Second attempts.



Test 12: Glasgow Voice Matching Test (GVMT)

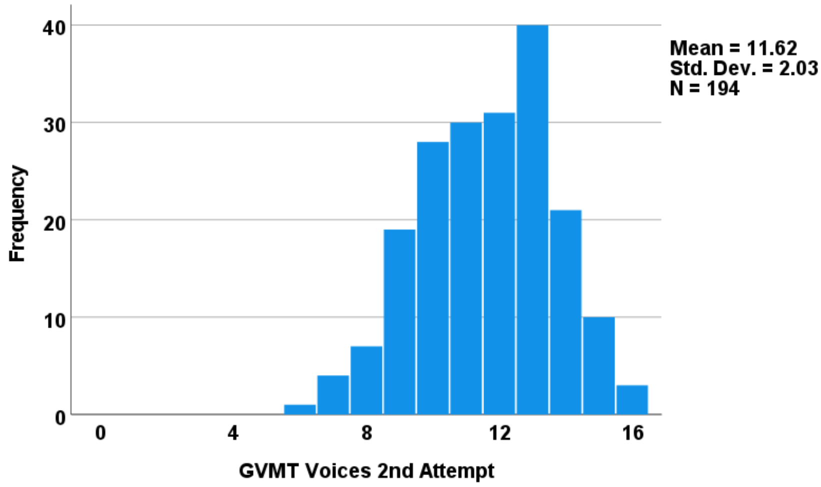
Figure 13.1: Glasgow Voice Memory Test (GVMT): First attempts.



GVMT "population norms" were compiled from the data reported in Aglieri et al. (2017)

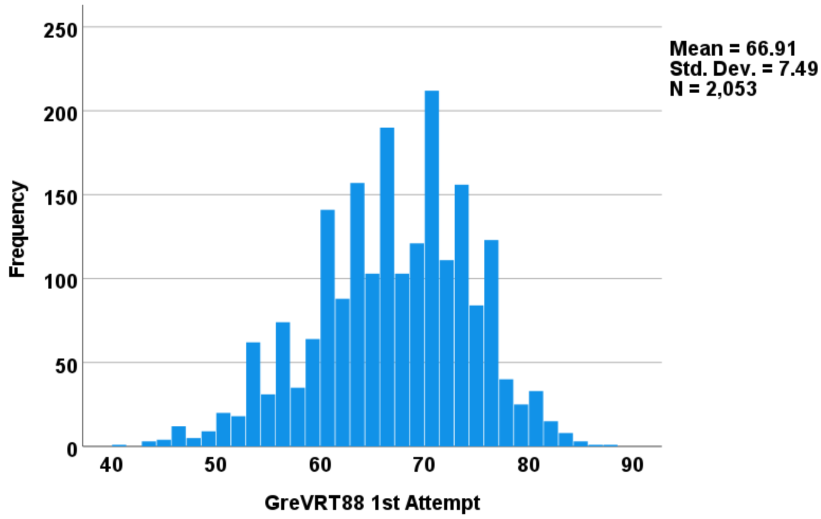
	Score	SD
μ (S)	12.50	(1.75)
z = -1.00	10.75	Lower limit for typical-ability controls
z = +1.00	14.25	Upper limit for typical-ability controls
z = +1.50	15.07	Liberal Super-Recogniser Threshold
z = +2.00	16.00	Strict Super-Recogniser Threshold
z = +2.00	16	Maximum score

Figure 13.2: Glasgow Voice Memory Test (GVMT): Second attempts.



Test 13: Greenwich Voice Memory Test 88: Version 1 (GreVMT88)

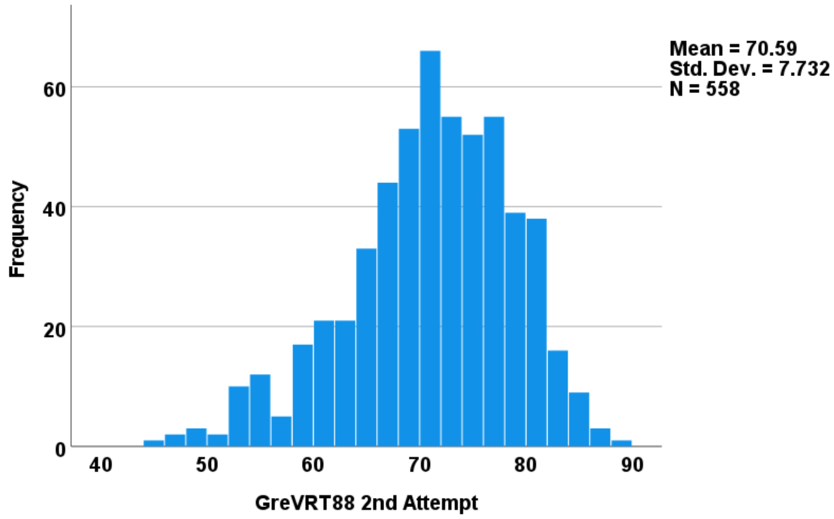
Figure 14.1: Greenwich Voice Memory Test 88 (GreVMT88): First attempts.



GreVRT88 “population norms” were compiled from the volunteer database, 21.04.2024 (n = 720)

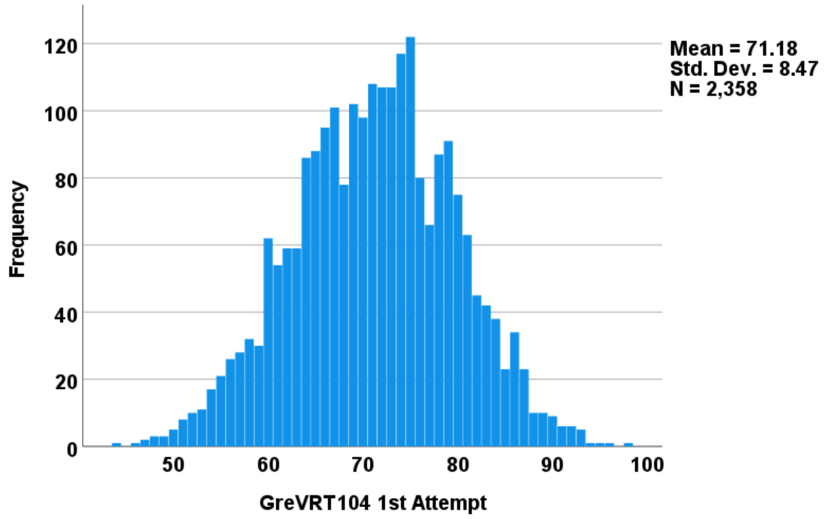
	Score	SD
μ (S)	66.62	(7.26) (z = 0)
z = - 1.00	59.36	Lower limit for typical-ability controls
z = +1.00	73.88	Upper limit for typical-ability controls
z = +1.50	77.51	Liberal Super-Recogniser Threshold
z = +2.00	81.14	Strict Super-Recogniser Threshold
z = +2.94	88.00	Maximum score

Figure 14.2: Greenwich Voice Memory Test 88 (GreVMT88): Second attempts.



Test 14: Greenwich Voice Memory Test 104 (GreVMT104)

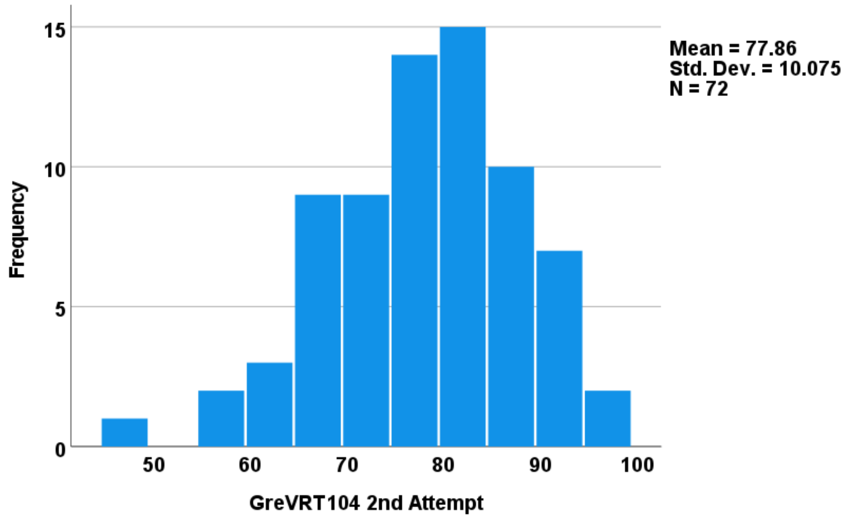
Figure 15.1: Greenwich Voice Memory Test 104 (GreVMT104): First attempts.



GreVRT104 “population norms” were compiled from the volunteer database, 21.04.2024 (n = 519)

	Score	SD
μ (S)	73.09	(8.24) (z = 0)
z = - 1.00	64.85	Lower limit for typical-ability controls
z = +1.00	81.33	Upper limit for typical-ability controls
z = +1.50	85.54	Liberal Super-Recogniser Threshold
z = +2.00	89.57	Strict Super-Recogniser Threshold
z = +3.75	104.00	Maximum score

Figure 15.2: Greenwich Voice Memory Test 104 (GreVMT104): Second attempts.



Test 15: Jena Voice Learning and Memory Test (JVLMT)

Figure 16.1: Jena Voice Learning and Memory Test (JVLMT): Proportion correct (Max = 22 trials)

JVLMT population norms were derived from the data reported in Humble et al. (2023)

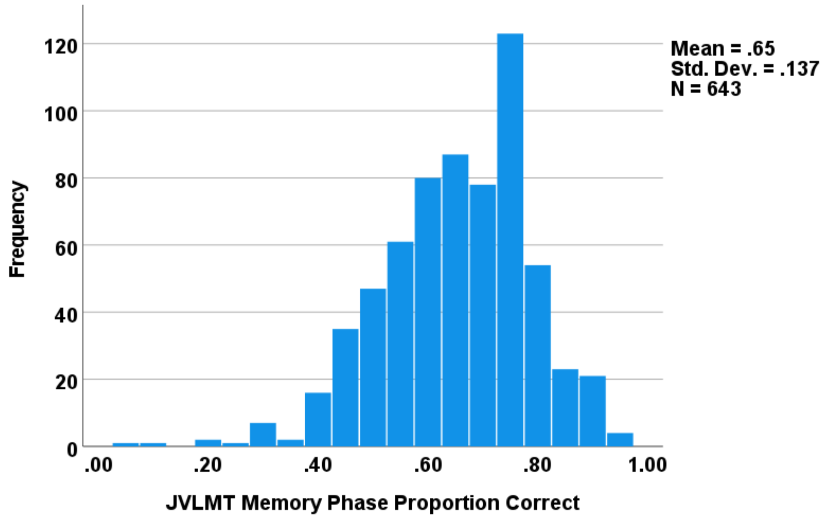
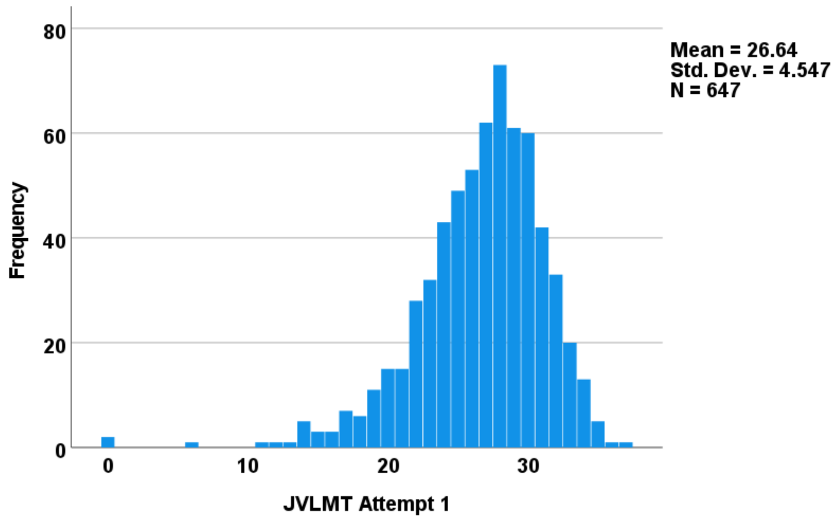


Figure 16.2: Jena Voice Learning and Memory Test (JVLMT): Score out of 38



Note: This JVLMT consists of a learning phase (16 trials) and a memory phase (22 trials). However, test's creators ignored the outcomes of the learning trials and they also excluded memory phase trials in which participants failed to make a response within 4 seconds. Test outcomes were reported as the proportion of items correctly responded to.

Two figures reporting the results of the test are included in this document. The first figure used the same criteria as the original article. However, to increase the sensitivity of the test, the University of Greenwich analysed the results of all 38 learning and memory trials, and instead of excluding trials which were not responded to within the time limits, these trials were treated as having been incorrectly responded to. The second figure reports the results expressed in terms of accuracy out of 38. Unusually a large proportion of participants who completed this test did not take sufficient numbers of the other voice tests for normative data to be produced. Therefore, criteria for super-recogniser and control groups are not reported.

Defining face recognition ability for research

We are certain that many people reading this blog will be hoping to find out if they can be defined as a super-recogniser based on their scores on these tests. Depending on the specific research project, the University of Greenwich and our collaborators have tended to define someone as a super-recogniser if they have achieved scores on 2-4 of our key tests that approximately 2% of a representative sample of the population would be expected to achieve. More formally, the 2% value is roughly the same as a score 2 standard deviations above the mean (z = 2.00). Therefore, our response is that if your scores reach the strict 2 standard deviations above the mean cut offs listed for most of the tests you have taken, then it is very likely that you do possess super-recognition levels of ability.

However, there is actually little consensus among researchers about these standards. Most use the 2% of the population standard. However, recently, there has been a tendency for researchers to create their own tests in order to measure face recognition ability, and this may be because the focus of the research tends to differ between research groups. However, this makes is virtually impossible to ensure roughly similar standards are maintained. This is one of the reasons we have asked for volunteers to take the UNSW Test, so we can measure our standards against the research group from the University of New South Wales (UNSW) in Sydney, Australia.

One thing that all researchers agree on is that increasing the number of tests used in a face recognition test battery increases the accuracy of face recognition ability estimates. However, in comparison to when a single test is used, lower proportions will achieve the 2 standard deviations above the mean cut offs on two tests and as the number of tests in a battery increases, participants achieving these strict super-recognition standards on all tests (regardless of the exact criteria chosen) become increasingly rare.

Six key tests of face recognition ability (in order of sensitivity)

- Cambridge (Male) Face Memory Test: Extended (CFMT+)
- Kent Face Matching Test (KFMT)
- Glasgow Face Matching Test: Version 2 (High 40) (GFMT2HI40)
- Short-Term Face Memory Test 30-60 (STFMT3060)
- Cambridge (Female) Face Memory Test (CFemFMT)
- Glasgow Face Matching Test (GFMT)

First or second test attempts? Normally, the first attempt on a test is the most reliable as everyone starts from an equal place. They will all be unfamiliar with procedures, and they have probably never seen any of the faces depicted in that test before. Most people improve with their second attempt on a test and yet, their face recognition ability will most likely remain at the level of their first attempt. Therefore, we normally use the first attempt score on each test in our estimates of face or voice recognition ability.

However, it would not be fair to retain the first test score if it was negatively impacted by problems that you had no control over. For instance, internet breakdowns, tiredness, illness, distractions, misunderstanding or misinterpreting instructions, and/or loss of concentration. For this reason, we have protocols for when we are informed of problems by participants or when we detect an anomaly when running automated checks, If we are informed of problems by participants, we normally suggest they make a second attempt on a score and we will use that instead, as long as that attempt is not made immediately, when the advantage from remembering the stimuli will be highest.

If a volunteer does not inform us of a problem, but If a first attempt test score is statistically far worse than the others in a volunteer's profile, this is often an indication of problems that were not identified at the time. We will automatically use the second attempt score instead. Therefore, we encourage participants to take all six tests at least twice, just in case we are waiting to replace a problematic first attempt score, with a second attempt effort. Note - you may take them as many times as you like. If you wish to do this, please take your time – there is no hurry – but, please rest between attempts. By doing this, you will help to ensure your ability estimates are as accurate as possible.

Appendix B: Tests described in this blog.

Could You be a Super-Recogniser Test (CYBSRT). This test is mostly provided for fun as no 14-trial test will be able to reliably measure the entire spectrum of face recognition ability in humans. Nevertheless, moderate positive correlations are normally found between scores on this test and other short-term face memory tests, and it therefore can be used as a reasonable predictor as to whether someone could possess good, typical-range, or poor face recognition ability. We mostly include it at the start of police projects as it provides an excellent introduction as to what to expect and participants are more practiced and reportedly less anxious, when they take the more reliable tests later.

It is almost certain that someone scoring below 10 out of 14 is not a super-recogniser. However, a high proportion of participants who take the test, use a mobile phone despite our advice not to. As the faces in the arrays may be too small on some phones to properly view key facial features, we have retained a message on completion, stating, "if you scored 10 or above, you could be a super-recogniser".

Cambridge Face Memory Test: Extended (CFMT+) (Russell et al., 2009). This 102-trial standardised short-term face memory test is probably the most commonly used worldwide test of face recognition, although it was not originally designed to measure super-recognition. Due to anomalies such as that the faces are mainly depicted with no hair, and that it has been available on the internet for participants to practice over a number of years, it would be unwise to define anyone as a police super-recogniser based on the results of this test alone. The Mean score on this test in most research is about 70 out of 102 (*SD* = 10), and scores of 90-95 out of 102 have been used to diagnose super-recognition in previous research, being representative of 2 SD above control means, a score likely to be achieved by about 2% of the population.

Glasgow Face Matching Test (GFMT) (Burton et al., 2010). This 40-trial test measures the ability to distinguish between two highly similar appearing white-ethnic facial images. It does not rely on memory. Participants decide whether 40 pairs of high-quality facial photographs depict the same person or not. Half of the trials are '*matched*' (i.e., the same person is depicted in the pair), half are *mismatch* trials. Participants are warned in advance of the randomly ordered but equal match-to-mismatch trial ratio. This test may be unsuitable for use to classify someone as a super-recogniser. Many participants score 100%. On the other hand, if a participant scores well under 40, they are very unlikely to be a super-recogniser.

Short-Term Face Memory Test 30-60 (STFMT3060). This 60-trial test measures short-term memory for unfamiliar black-and-white ethnic faces. In the learning phase of this test, 30 male faces are sequentially presented in identical purple sweatshirts for 10 sec. In the test phase, new photos of the 30 'old' faces are randomly intermixed with 30 'new' faces, wearing a variety of different sweatshirts. Participants respond as to whether faces are 'old' or 'new'.

The Kent Face Matching Test (KFMT) (Fysh & Bindemann, 2018). This 40-trial test measures the ability to distinguish between two highly similar appearing white-ethnic facial images. It does not rely on memory. Participants decide whether 40 pairs of high-quality facial photographs depict the same person or not. Half of the trials are '*matched*' (i.e., the same person is depicted in the pair), half are *mismatch* trials. Participants are warned in advance of the randomly ordered but equal match-to-mismatch trial ratio.

Cambridge Cars Memory Test (CCarMT) (Dennet et al., 2011). This 72-trial standardised short-term memory test has an identical structure to the short-version of the Cambridge Face Memory Test. It allows us to extract cars from face scores to generate an estimate of Non-Face Object Memory Ability. However, the ability to recognise cars is partly based on exposure and interest in cars. We ask a question around this, to ensure we can interpret the results appropriately. However, an aim for 2022-2023 is to make available a series of three very different object memory tests. We think it unlikely anyone will be an expert on all four tests (including cars).

Cambridge (Female) Face Memory Test (CFemFMT) (Arrington et al., 2021). This 102-trial test was designed to act as a female equivalent for the Cambridge (Male) Face Memory Test: Extended (CFMT+). The authors demonstrated that mean scores on this test are approximately 13 out of 102 higher than mean scores on the male version of the test.

Glasgow Face Matching Test: Version 2 (High 40) (GFMT2HI40) (White et al., 2022). This 40-trial test is an update of Version 1 of the *Glasgow Face Matching Test*, using images taken from the same database. The authors produced different versions of the test to measure participants of different abilities. We use the High 40 version, which was designed for use with participants possessing superior ability.

Jena Voice Learning and Memory Test (JVLMT) (Humble et al., 2023). This test consists of 16 learning trials and 22 test trials. The authors employed a scoring system based on the 22 test trials. However, we have used the full set of 38 trials to increase sensitivity.

List of References

Aglieri, V., Watson, R., Pernet, C., Latinus, M., Garrido, L., & Belin, P. (2017). The Glasgow Voice Memory Test: Assessing the ability to memorize and recognize unfamiliar voices. *Behavior Research Methods*, *49*, 97-110. https://doi.org/10.3758/s13428-015-0689-6

Anastasi, J. S., & Rhodes, M. G. (2005). An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin & Review, 12,* 1043-1047. DOI: 10.3758/BF03206441

Arrington, M., Elbich, D., Dai, J., Duchaine, B., & Scher, K. S. (2022). Introducing the female Cambridge face memory test – long form (F-CFMT+). *Behavioral Research, 54,* 3071–3084 (2022). https://doi.org/10.3758/s13428-022-01805-8

BBC Future (11 June 2015). The super power police now use to tackle crime. Venkataramanan, M. https://www.bbc.com/future/article/20150611-the-superpower-police-now-use-to-tackle-crime

Belanova, E., Davis, J. P., & Thompson, T. (2018). Cognitive and neural markers of super-recognisers' face processing superiority and enhanced cross-age effect. *Cortex*, *98*, 91-101. DOI:10.1016/j.cortex.2018.07.008 (Download pre-print here: https://bit.ly/bdtb2018)

Bobak, A. K., Pampoulov, P., & Bate, S. (2016). Detecting superior face recognition skills in a large sample of young British adults. *Frontiers in Psychology*, *7*(1378). https://doi.org/10.3389/fpsyg.2016.01378

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods*, *42*(1), 286-291. DOI:10.3758/BRM.42.1.286

Business Insider Australia (29 October 2015). A select group of people have an uncanny ability to remember faces for years. Lewis, T. http://www.businessinsider.com.au/people-who-are-good-at-remembering-faces-2015-10

Correll, J., Ma., D. S., & Davis, J. P. (2020). Perceptual tuning through contact? Contact interacts with perceptual (not memory-based) face-processing ability to predict cross-race recognition. *Journal of Experimental Social Psychology, 92*, 104058, https://doi.org/10.1016/j.jesp.2020.104058

Daily Mail (15 June 2015). Are YOU a ‘super recogniser’? Take the test to see if you are one of an elite group of people who never forget. Zolfagharifard, Z. http://www.dailymail.co.uk/sciencetech/article-3125173/Are-super-recogniser-test-one-elite-group-people-rarely-forgets-face.html

Daily Mirror (15 June 2015). Are you a super recogniser? Take the test Met police use to help find criminals. McCrum, K. http://www.mirror.co.uk/news/weird-news/you-super-recogniser-take-test-5892509

Davis, J. P. (2013). Super-recognisers in the Metropolitan Police. *IFSEC Global*, 15 October 2013 (download pre-print here)

Davis, J. P. (2020). CCTV and the super-recognisers. In C. Stott, B. Bradford, M. Radburn, and L. Savigar-Shaw (Eds.), Making an Impact on Policing and Crime: Psychological Research, Policy and Practice (pp 34-67). London: Routledge. ISBN 9780815353577. https://doi.org/10.4324/9780429326592 (Download free pre-print here: https://bit.ly/34Phwjm)

Davis, J. P., Bretfelean, D., Belanova, E., & Thompson, T. (2020). Super-recognisers: face recognition performance after variable delay intervals. *Applied Cognitive Psychology*, *34*(6), 1350-1368. DOI:10.1002/acp.3712 (Download free pre-print here: https://bit.ly/3slkg0m)

Davis, J. P., Lander, K., Evans, R., & Jansari, A. (2016). Investigating predictors of superior face recognition ability in police super-recognisers. *Applied Cognitive Psychology, 30*(6), 827–840. DOI: 10.1002/acp.3260 (Download pre-print here: https://bit.ly/dlej2016)

Davis, J. P., Lander, K., & Jansari, A. (2013). I never forget a face. T*he Psychologist, 26*(10), 726-729. (read here - open access)

Davis, J. P., Treml. F., Forrest, C., & Jansari, A (2018). Identification from CCTV: Assessing police super recognisers ability to spot faces in a crowd and susceptibility to change blindness.

Applied Cognitive Psychology, 32(3), 337-353. DOI: 10.1002/acp.3405 (Download pre-print here: https://bit.ly/3dtfj2018

Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., & Duchaine, B. (2011). The Cambridge Car Memory Test: A task matched in format to the Cambridge Face Memory Test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects. *Behavior Research Methods*, *44*(2), 587–605. https://doi.org/10.3758/s13428-011-0160-2

Foster, J. H., Martin, C., & Davis, J. P. (2021). Alcohol consumption during the Covid lockdown period: Predictors of at-risk drinking at different AUDIT-C cut-off thresholds. I*nternational Journal of Environmental Research and Public Health: Health Behavior, Chronic Disease and Health Promotion*, (19), 42, 13042. https://doi.org/10.3390/ijerph182413042 (free open access)

Fysh, M. C. (2018). Individual differences in the detection, matching and memory of faces. *Cognitive Research: Principles and Implications*, *3*(1), 1-12. https://cognitiveresearchjournal.springeropen.com/articles/10.1186/s41235-018-0111-x

Fysh, M. C., & Bindemann, M. (2018). The Kent face matching test. *British Journal of Psychology*, *109*(2), 219-231. https://doi.org/10.1111/bjop.12260

Gentry, N. W., & Bindemann, M. (2019). Examples improve facial identity comparison. *Journal of Applied Research in Memory and Cognition*, *8*(3), 376-385.

Gov.uk (2020). Population of England and Wales. Downloaded 20 November 2021 from, https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/population-of-england-and-wales/latest

Herlitz, A., & Lovén, J. (2013). Sex differences and the own-gender bias in face recognition: A meta-analytic review. *Visual Cognition*, *21*(9-10), 1306-1336. https://doi.org/10.1080/13506285.2013.823140

Humble, D., Schweinberger, S. R., Mayer, A., Jesgarzewsky, T. L., Dobel, C., & Zäske, R. (2023). The Jena Voice Learning and Memory Test (JVLMT): A standardized tool for assessing the ability to learn and recognize voices. *Behavior Research Methods*, *55*(3), 1352-1371. https://doi.org/10.3758/s13428-022-01818-3

LASIE Partners (2014-2017). LArge scale information exploitation of forensic data (LASIE). *European Commission 7th Framework Programme. SEC-2013.1.6-1: 607480.* https://cordis.europa.eu/project/id/607480

Mcintosh, S., & Davis, J. P. (2020). The 'casting couch' scenario: Impact of perceived employment benefit, reporting delay, complainant gender, and participant gender on juror decision-making in rape cases. *Journal of Interpersonal Violence*, DOI: 10.1177/0886260520966679, Download pre-print here: http://bit.ly/McintoshJPD

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*(1), 3-35. https://doi.org/10.1037/1076-8971.7.1.3

Mühl, C., Sheil, O., Jarutytė, L., & Bestelmeyer, P. E. (2018). The Bangor Voice Matching Test: A standardized test for the assessment of voice perception ability. *Behavior Research Methods*, *50*, 2184-2192. https://doi.org/10.3758/s13428-017-0985-4

New York Times (9 October, 2015). London Police 'Super Recognizer' Walks Beat With a Facebook of the Mind. Bennhold, K. http://www.nytimes.com/2015/10/10/world/europe/london-police-super-recognizer-walks-beat-with-a-facebook-of-the-mind.html?_r=0

Robertson, D., Black, J., Chamberlain, B., Megreya, A. M., & Davis, J. P. (2020). Super-recognisers show an advantage for other race face identification. *Applied Cognitive Psychology*, *34*(1), 205-216. DOI: 10.1002/acp.3608 (Download free pre-print here: https://bit.ly/rbcmd2020)

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252–257. https://doi.org/10.3758/PBR.16.2.252

Satchell, L., Davis, J. P., Julle-Danière, E., Tupper, N., & Marshman, P. (2019). Recognising faces but not traits: Accurate personality judgment from faces is unrelated to superior face memory. *Journal of Research in Personality*, https://doi.org/10.1016/j.jrp.2019.02.002 (Download pre-print here: https://bit.ly/sdjdtm2019)

South China Morning Post (25 October 2015). British academic launches search for 'super recognisers' in Hong Kong who never forget a face. Lam, L. http://www.scmp.com/news/hong-kong/education-community/article/1871808/british-academic-launches-search-super

White, D., Guilbert, D., Varela, V.P.L., Jenkins, R., & Burton, A. M. (2022). GFMT2: A psychometric measure of face matching ability. *Behavioral Research, 54*, 252–260 (2022). https://doi.org/10.3758/s13428-021-01638-x