Scores of high range tests' candidates on standard psychometric batteries & tests by Dr. Jouve.

hriqtests.com, July 2019

Aim of this document is to provide information regarding the "statistical profile" of high range tests' candidates, according to their reported performance on standard psychometric batteries & tests created by Dr. Xavier Jouve.
Although consisting of mostly simple, descriptive, statistics, it's more than valuable in drawing certain conclusions regarding mean (and expected) performance on - and norming of - high range tests, as some of them may suffer from inflated/deflated scores and insufficient statistical processing. Plain language is used, in order this report to be approachable to anyone interested, regardless special knowledge on statistics.
At this point, it has to be mentioned that tests of Dr. X. Jouve were chosen due to the fact that they have been normed using quite large testing samples (most of them on N>300, consisting mainly of high achieving individuals with mean score >131, M=100, SD=15 on standard psychometric batteries).
In addition, one may notice that both r (Pearson correlation coefficient) and ρ (Spearman's rho, Spearman's rank correlation coefficient) are presented at Part 5 of this report. Although there are certain indications that scores of high-range tests' candidates distribute almost "normally", a non-parametric measure as the Spearman's Rho (ρ) will be more reliable until it's proven otherwise.
Histograms, some parameters and further discussion, follow.
In addition, an overall correlation of theoretical IQs provided by tests on hriqtests.com with standard psychometric batteries and tests by X. Jouve is provided.

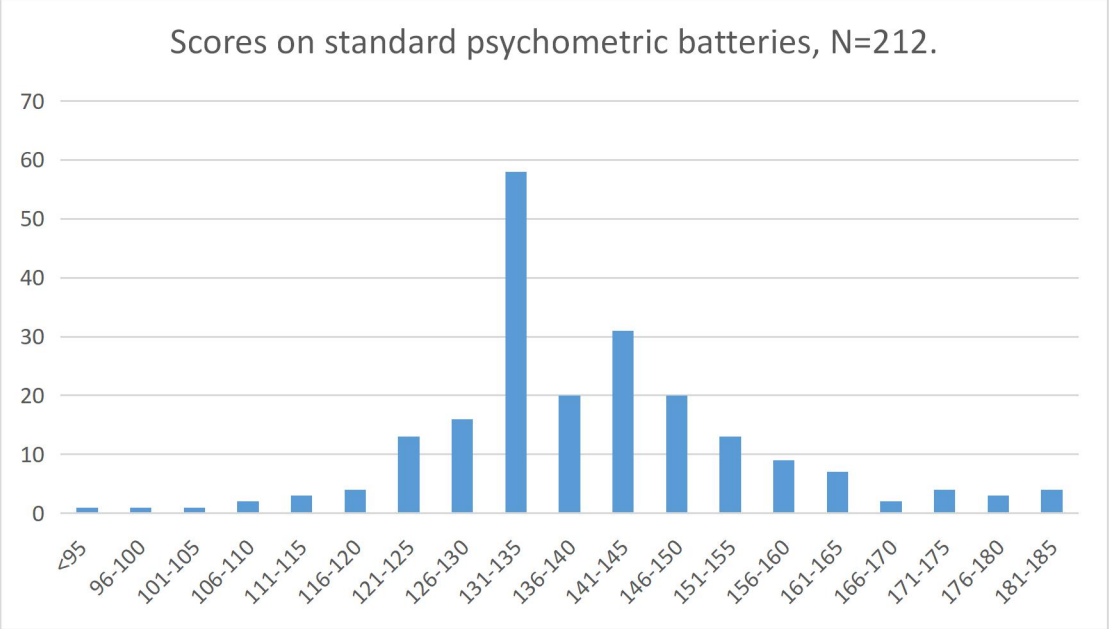
Tests used (Scores' N=350) :
A) Standard psychometric batteries (alphabetically): BADyG-M (2), BLS4-2T (1), CCFIT III (17), CFT20-R (2), FRT-A (12), FRT-B (9), GAMA (1), IBF-S (3), IDF (1), IST 2000R (2), IST 70 (3), MAT (10), NGCT (1), OLSAT (2), PCAT (1), RAIT (1), RAPM (27), RIAS (1), RSPM (5), SBIS (5), TOGRA (1), WAIS (R,III,IV - 70), WISC (9), WPT (4), Unknown Mensa Entrance Test (Probably RAPM or FRT - 21).

B) Tests by Dr. X. Jouve (alphabetically): C-09 (17), C-10 (14), C-12 (1), CCAT (2), CSE (abbreviated NVCPE-R form, granted after C-09 to high-scoring contestants - 3), IAW (2), JCCES (9), JCTI (23), NVCPE (3), NVCPE-R (59), TLAP-R (3), TRI (2).

Parts
1. Scores on standard psychometric batteries (N=212).
2. Scores on standard psychometric batteries (N=172, low-ceiling & extrapolated scores excluded).
3. Scores on Dr. Jouve's tests (N=138).
4. Scores on standard psychometric batteries & Dr. Jouve's tests (N=310).
5. Overall correlation of hriqtests.com tests with standard psychometric batteries & Dr. Jouve's tests.

(Proceed to next page)

1. Scores on standard psychometric batteries.



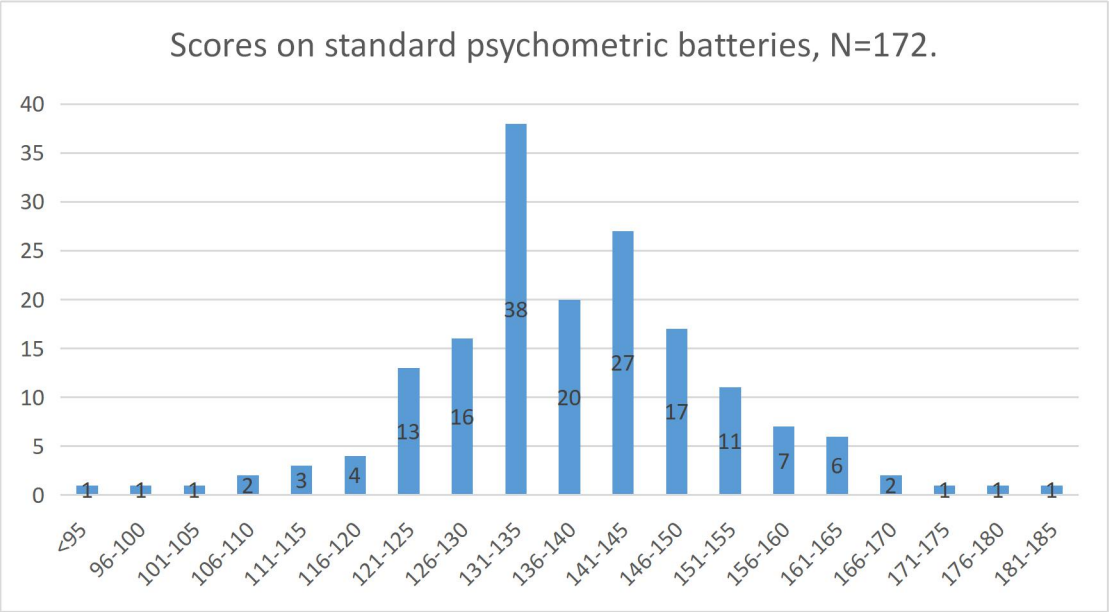
All reported scores are included (fixed scores, low-ceiling scores (scores >x are presented as x), extrapolated scores). Lowest score is 85 sd15 (1, WAIS) and highest scores are 185 sd15 (2 extrapolated, WAIS & SBIS).

Mean Score : 140,34.
Median : 137,5.
Standard Deviation : 15,23.

As a lot of extrapolated and ceiling scores are included (N=40), no significant conclusions can be made.

(Proceed to next page)

2. Scores on standard psychometric batteries (low-ceiling & extrapolated scores excluded).



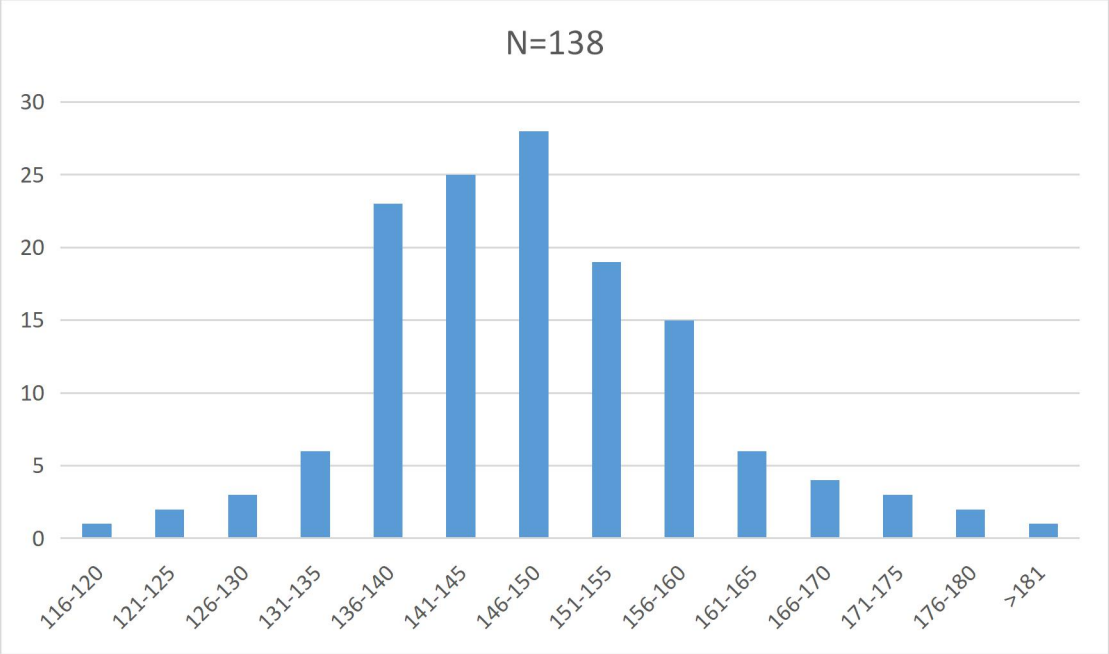
Lowest score is 85 sd15 (1, WAIS) and highest is 182 sd15 (1, CCFIT III).

Mean Score : 138,63.
Median : 137,5.
Standard Deviation : 14,04.

Distribution of scores is quite close to normal. A possible anomaly is easily noticed at 131-140 range and there are certain reasons for this. According to my experience and in my humble opinion, it's indeed an anomaly occurring at the aforementioned range and we are not talking about a distribution different than the normal one; there is a certain reason for this : A lot of assessment tools (like FRTs) provide scores up to that level or close to that level. So, there might be either not that reliable results in case of people reaching a test's ceiling or some ceiling scores may have unwittingly been reported as fixed ones.
As one may notice in one of the next histograms, in which scores on standard psychometric batteries are combined with the ones on tests by Dr. Jouve, such an anomaly is not present. And do notice that Dr. Jouve's tests have significantly higher ceilings than 131-140.

(Proceed to next page)

3. Scores on tests by Dr. Jouve

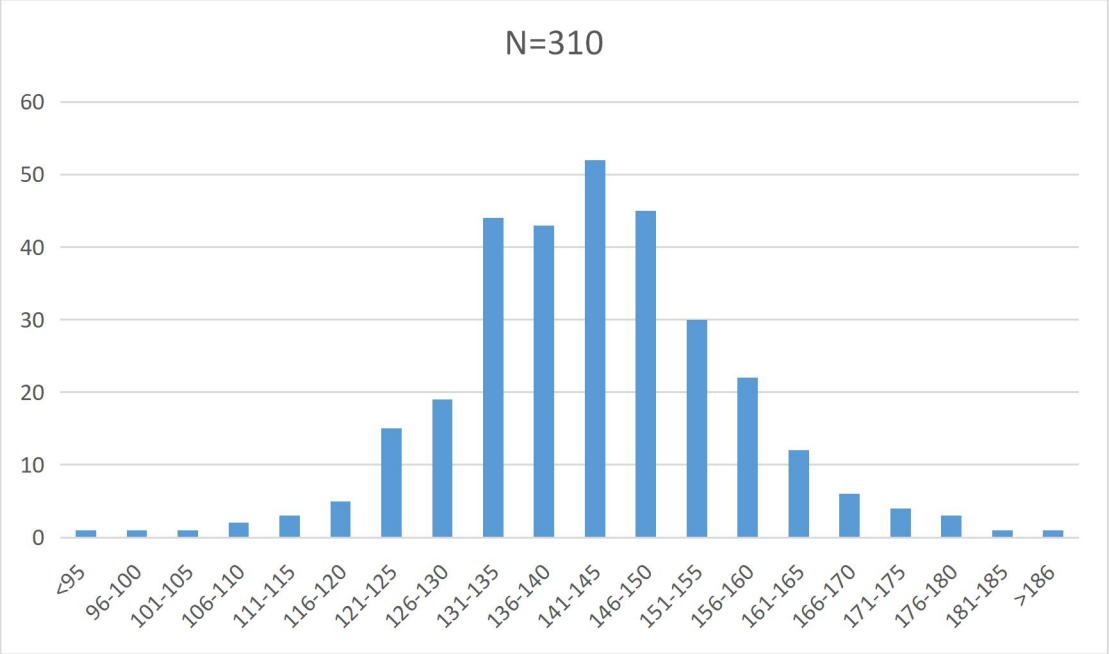


Lowest score is 120 sd15 (1, JCTI) and highest 198 sd15 (1, NVCPE-R, Rasch-analysis equated. Fluid Intelligence Index was reported as 181 sd15). That was the highest score ever reported, until quite recently, when a score of 204 sd15 was announced.

Mean score : 148,02.
Median : 147.
Standard Deviation : 11,45.

(Proceed to next page)

4. Scores on standard psychometric batteries & tests by Dr. Jouve.



Lowest score is 85 sd15 (1, WAIS) and highest 198 sd15 (1, NVCPE-R).

Mean score : 142,81.
Median : 142.
Standard Deviation : 13,75.

(Proceed to next page)

5. Overall correlation of hriqtests.com tests with standard psychometric batteries & Dr. Jouve's tests.

Correlations between theoretical IQs provided (until late July, 2019) on hriqtests.com's tests (CPE-N, VRA50, The Phi Test, SEE30) and : a) Standard psychometric batteries (Pairs' N=99, low-ceiling and extrapolated scores excluded), b) Dr. Jouve's tests (Pairs' N=125), c) Combination of the aforementioned (N=224), are provided.

r: Pearson correlation coefficient (Pearson's r).
ρ: Spearman's rank correlation coefficient (Spearman's rho)
MIQ : Mean IQ on the mentioned test(s) (SD=15).
MIQ-HR : Mean theoretical IQ provided by tests of hriqtests.com (SD=15).

Test	Pairs (IQ range)	r	ρ	MIQ	MIQ-HR	P-Value
Total	224 (85-172)	0.73	0.77	140.72	141.03	<0.001
Supervised	99 (85-172)	0.74	0.79	136.82	137.53	<0.001
Jouve	125 (120-170)	0.71	0.73	143.8	143.81	<0.001
WAIS	32 (85-164)	0.86	0.84	135.66	135.38	<0.001
RAPM	20 (111-147)	0.52	0.67	133.6	136.15	0.0012
JCTI	34 (123-160)	0.69	0.67	143.29	144.91	<0.001
C-09	31 (120-153)	0.8	0.82	140.18	142.84	<0.001

Correlation of 0.7+ with a significant amount of scores on standard psychometric batteries and especially that of 0.8+ with WAIS, that is the most used and up-to-date intelligence assessment tool, shows that the work done in the past (tests' revisions, certain kind of items selection according to discrimination indexes etc.), was of utmost importance and had to be done. New generation of tests (CPE-N, SEE30, VRA50, The Phi Test), as well as the currently running contests (INSC & IVAC) have proven themselves as ones of high quality.
It's also proven that common work is needed, between authors and tests' candidates. Tests' authors should thoroughly examine their tests statistically and DEFINITELY revise them if needed (it is, until they reach some certain level of quality) and candidates should report other scores (and, by this way, help in author's statistical "journey").
To sum up, I will repeat something written at a former report: "Regarding test designing, after 8 years of having created several of them, one thing can be said for sure: There exist specific (and different, of several difficulty levels) patterns (in other words, kinds of items) that tend to behave steadily and "healthy" – that is, keeping their difficulty regardless the environment (whether they are part of one test or another), showing high discrimination index. It's only after 8 years of studying and experimenting with tests that I can now say that I 've created a team of decent tests. And in this point, one should realize that "attractiveness" or "originality" of a test do not necessarily guarantee its quality. Of course, combination of the aforementioned is the ultimate goal."