



The
Crucible

THE PRIVATE EQUITY
HORSEPOWER ASSESSMENT®

Crucible Insights



Population Variance Testing

2023



WHY IS POPULATION VARIANCE TESTING IMPORTANT?

Population variance testing provides insight into how data points in specific populations are spread out. These insights are used to adjust for existing bias, which diminishes the validity of an assessment. Bias is a systematic error introduced into sampling or testing by selecting or encouraging one outcome or answer over others.¹ The Crucible is a leadership research tool designed specifically for middle market private equity firms, their portfolio companies, and similar high stakes environments.² To accurately assess individuals and engage the best leaders, assessments must be free of bias that would reduce potential diversity in portfolio companies. By testing for potential bias, this demonstrates that The Crucible provides a fair, and high quality evaluation for all leaders.

After thorough analysis, The Crucible is fully prepared to evaluate leaders of all backgrounds and identities for their fit in these environments.

SAMPLING PROCEDURE

To collect the data, The Crucible partnered with Qualtrics, the most trusted research software, to conduct a general population survey. Qualtrics is used by 85% of Fortune 100 companies and conducts over a billion surveys each year.³

An important part of defining the sample is determining the appropriate sample size. Four factors were evaluated to determine sample size: population size, margin of error, confidence level, and standard deviation. These factors are used to calculate a z-score, which is used to calculate sample size.⁴

$$\text{Calculated sample size} = \frac{Z^2 * SD * (1-SD)}{(\text{margin of error})^2}$$

A sample size of 305 was determined with a pre-established confidence level of 95% which is commonly used with the analyses conducted in this testing effort. Additionally, quotas were added to tailor the sample towards those most likely to be taking The Crucible. Respondents were required to be 30 years of age or older, have obtained at least a Bachelor's degree, and have previously held a management position. Respondents were also removed if they did not spend a reasonable amount of time completing the assessment (speeding) or if their response patterns were clearly inattentive (selecting only the first answer for every question).

Guidelines for general US population age brackets, gender, and race quotas were added to ensure the groups were large enough to detect significant differences across all identities.

Gender Male - ~30% Female - ~70%	Race Non-Hispanic White - ~30% Non-Hispanic Black - ~23% Hispanic - ~23% Other Races - ~23%
---	--

¹ <http://www.merriam-webster.com/dictionary/bias>.

² <https://thecrucible.com/>

³ <https://www.qualtrics.com/core-xm/survey-software/Qualtrics>

⁴ <https://www.qualtrics.com/experience-management/research/determine-sample-size/>

SAMPLE

The following tables define the sample of respondents used to complete population variance testing including the gender, race/ethnicity, and age of respondents.

Total Sample Population (N=305)

Gender	Number in Sample	Percent in Sample
Male	92	30.16%
Female	213	69.84%

Race	Number in Sample	Percent in Sample
Asian	66	21.64%
Black or African American	79	25.90%
Hispanic or Latinx	65	21.31%
Middle Eastern or North African	3	0.98%
Native American	3	0.98%
Native Hawaiian or Pacific Islander	1	0.33%
White	103	33.77%

Age	Number in Sample	Percent in Sample
<30	0	0.00%
30-40	164	53.77%
40-50	110	36.07%
50-60	19	6.23%
60-70	12	3.93%
70+	0	0.00%

Male (N=92)

Race	Number in Male Sample	Percent in Male Sample
Asian	9	10%

Black or African American	16	17%
Hispanic or Latinx	9	10%
Middle Eastern or North African	0	0%
Native American	0	0%
Native Hawaiian or Pacific Islander	0	0%
White	60	65%

Age	Number in Male Sample	Percent in Male Sample
<30	0	0%
30-40	61	66%
40-50	29	32%
50-60	1	1%
60-70	1	1%
70+	0	0%

Female (N=213)

Race	Number in Female Sample	Percent in Female Sample
Asian	57	27%
Black or African American	63	30%
Hispanic or Latinx	56	26%
Middle Eastern or North African	3	1%
Native American	3	1%
Native Hawaiian or Pacific Islander	1	0%
White	43	20%

Age	Number in Female Sample	Percent in Female Sample
<30	0	0%
30-40	103	48%

40-50	81	38%
50-60	18	8%
60-70	11	5%
70+	0	0%

THE CRUCIBLE REPORT AND THE POTENTIAL AREAS OF VARIANCE:

The Crucible is a leadership research tool. It is designed to identify and evaluate characteristics pertinent to thriving in high stakes environments. The assessment provides a PEQ (Private Equity Quotient) score, an Archetype which pinpoints an individuals' transformation bias and leadership style, and Five Catalysts that highlight the most important attributes in value creation in the private equity backed middle market (and elements within those catalysts). The report also offers insights to certain Contaminants that are detrimental in this environment. Contaminants are used to provide insight in areas for development. Averages of the entire dataset are provided.

Beginning at the element level and working up through the PEQ, no bias was identified in the analysis. The Contaminants were also tested and analyzed. The following table breaks down all areas tested for potential biases.

Catalysts: Elements	Execution Engine: Urgency, Relentlessness, Impatience, Hands-on-Approach, Digs for Root Cause, Detail Orientation, Systems and Process Focused, Metrics Focused, Intellectual Curiosity	Emotional Leverage: Passion, Intensity, Hunger/Drive, Resiliency, Energy/Endurance, Disdain for Losing, Competitiveness,	Leadership: Sets High Standards, Optimism, Inspires Confidence and Belief, Directness, Strategic Perspective, Ensures Buy-in, Engages Deeply with Stakeholders, Drives Accountability	Cognitive Horsepower: Logic, Math, Word Association, Pattern Recognition	Interpersonal: Self Esteem, Belief in Others, Empathy, Confidence, Authenticity, EQ, Collegiality
----------------------------	---	--	---	--	---

Contaminants: Hesitancy, Fear of Failure, Inability to Adapt, Lack of Awareness, Fragility, Arrogance

STATISTICAL ANALYSIS:

The Crucible undertook an initial statistical analysis, looking at the assessment in the broadest view.

Examining for bias initially at the highest and most inclusive level, gives the statistician the ability to ascertain general areas of bias before conducting a deep dive into the more granular aspects of the survey. It is important to collect this data to help facilitate further statistical analysis, and to have a better understanding of the data set itself.

The preliminary testing utilized a Chi Squared Test to get understanding of overall potential areas of bias between the Catalysts.

$$\text{Chi Squared Test } \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

A Chi Squared Analysis is a test that measures how a model compares to actual observed data. The Chi Square statistic compares the size of any discrepancies between the expected results and the actual results, given the size of the sample and the number of variables in the relationship.⁵ The test indicated potential bias within the Interpersonal Catalyst. This potential for bias necessitated further testing to identify the specific Elements within the Catalyst and other nuances that might have created a statistically significant difference.

A Chi Squared test allows a researcher to say that either “we can reject the null hypothesis of no relationship at the 0.05 level or “we have insufficient evidence to reject the null at the 0.05 level.”⁶ The relationship, therefore, that a Chi Squared test identifies is not sufficient to test the individual elements for bias. The Chi Squared test requires the use of data that is independent. As each Element score represents several question scores taken from the same person, the data are not independent. The correct number of responses an individual gets on the first ten questions probably is correlated with the number of correct responses on the second ten questions. As the data are potentially correlated, and not independent, the chi square test cannot be used appropriately. Therefore, to analyze the elements, another statistical test was needed: a T-test.

A T-test is one of the most fundamental tests for statisticians in comparing the mean of two given samples.

$$\text{T-Test } t = \frac{m-u}{s/\sqrt{n}}$$

A T-Test is a type of inferential statistic used to determine if there is a significant difference between the means of two groups. The tests were used to determine if the significance level was below 0.05. Two different types of T-tests were run to determine if bias existed. A one tail T-Test and a two tail T-Test. A one tail test determines if the sample mean would be higher or lower than the population mean, but not both. The two tailed tests both below and above the mean in 2.5% in both directions. A Welch’s t-test was used to account for the fact that group sizes and variances were unequal.⁷

SUMMARY OF FINDINGS:

A thorough statistical analysis of all scores at the element level, catalyst level, and finally the overall PEQ score, demonstrates no apparent bias in the assessment.

⁵ <https://www.investopedia.com/terms/c/chi-square-statistic.asp>

⁶ <https://sciencing.com/difference-between-ttest-chi-square-8225095.html>

⁷ <https://www.statology.org/welchs-t-test/>

A preliminary way to determine potential areas of bias amongst the Catalysts is to use a t-test to compare group mean differences. This test is used to determine if there is a significant difference between the average scores of two different groups, determined by a p-value below 0.05. A two-tailed test was used as it does not assume any directionality in group differences and is thus more appropriate. Welch's t-test was used because it is more appropriate when group sizes and variances are unequal.⁸

$$t = (X_1 - X_2) / \sqrt{(s_1^2/n_1) + (s_2^2/n_2)}$$

A cohesive summary including the t-tests comparing mean scores across all parts of the assessment for subgroups of interest can be found in the appendix. Here we note any statistically significant effects that were detected.

At the Catalyst level, we identified one area for exploration. There were statistically significant differences detected for the Interpersonal Catalyst such that female respondents scored higher than male respondents ($p = .036$). Also for the Interpersonal Catalyst, Black respondents scored higher than White respondents ($p = .048$). The details of these tests can be found in the appendix.

To explore these further we looked at Element level comparisons within the Interpersonal Catalyst. For the comparison of female to male respondents, female respondents scored statistically significantly higher than male respondents on two Elements - Authenticity ($p = .005$) and Empathy ($p = .003$). For the comparison of Black and White respondents, Black respondents scored statistically significantly higher than White respondents on two different Elements - EQ ($p = .009$) and Collegiality ($p = .050$). The Crucible is not concerned about negative impact being introduced by these significant differences.

While these tests show differences in mean scores between the tested groups, mean differences are not indicative of potential biases or adverse impact.⁹ When mean differences are detected, this can reflect true underlying differences in the overall populations, and this explanation is more likely when supported by other empirical research. The Crucible will continue to monitor and conduct follow-up studies.

It is a well-established finding that women score higher than men on integrity tests¹⁰ and measures of emotional intelligence.¹¹ These are similar to what The Crucible captures in the Interpersonal Catalyst, specifically Authenticity and Empathy. According to the research women generally have more complex emotional knowledge than men and are better at perceiving emotional cues. Furthermore, women who have not yet made it to leadership positions are often disadvantaged due to societal gender roles, however women who have emerged in top leadership positions often have a leadership advantage because they are seen as successful agentic and communal leaders¹²

Thus, the difference we detected is likely to be a reflection of real differences between men and women in interpersonal abilities rather than a function of test bias.

Research has also supported the finding that minority racial groups perform better on self-reported emotional intelligence measures, suggesting a similar heightened interpersonal ability.¹³

⁸ <https://www.statology.org/welchs-t-test/>

⁹ https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf (Pg. 65, Standard 3.6)

¹⁰ Berry et al. (2007); <https://doi.org/10.1111/j.1744-6570.2007.00074.x>

¹¹ Newman et al. (2010); <https://shrm.org/hr-today/news/hr-magazine/Documents/Joseph-Newman-2010.pdf>

¹² <https://msbfile03.usc.edu/digitalmeasures/tost/intellcont/Rosette%20and%20Tost%202010%20JAP-1.pdf>

¹³ Newman et al. (2010); <https://shrm.org/hr-today/news/hr-magazine/Documents/Joseph-Newman-2010.pdf>

Gender and racial differences can be further explained by the phenomena known as intersectionality, which refers to the way that a person's many social identities such as race, gender, sexuality, and socioeconomic status interact with one another to create unique experiences of privilege or oppression. Those who identify with more than one minority identity often face more interpersonal challenges and these experiences allow them to better connect with and empathize with others, especially in times of adversity¹⁴, thus explaining higher scores in EQ and collegiality.

Mathematically, increased minority representation in the sample could be causing additional differences to be detected. That is, if women have a real difference in interpersonal ability, we expect to find a difference across racial groups as well because 80% of the Black identifying respondents in our sample also identify as female. The majority of female respondents in our sample identified as non-white, while 65% of the male respondents identified as white. Increased minority representation in this sample was intended to ensure that minority groups were not disadvantaged by the Crucible assessment.

Importantly, these differences do not cause concern for adverse impact. The EEOC defines adverse impact as “a substantially different rate of selection in hiring, promotion, or other employment decisions which work to the disadvantage of members of a race, sex, or ethnic group.”¹⁵ To test this, a user of The Crucible may adopt a practical measure taken directly from The Uniform Guidelines on Employee Selection Procedures known as the 4/5ths or 80% rule.¹⁶ The rule states that adverse impact occurs when any group does not have a selection or passing rate equal to or greater than 80% of the group with the highest rate (most often white men). The Crucible, as a research tool, does not have a defined passing rate, so it cannot violate the 4/5ths rule absent the specific application made by a test user. Furthermore, because small mean differences found in this investigation are in favor of minority groups and the population of private equity leaders is disproportionately represented by the white male population,¹⁷ it is very unlikely that any calculated selection ratio using the Crucible scores could violate the 4/5ths rule. The selection ratio would have to be higher for a minority group that is underrepresented in private equity to demonstrate adverse impact based on these differences. Thus, when used in a specific application, there is no potential for adverse impact to occur when using The Crucible.

¹⁴ Lim & DeSteno (2016); <https://psycnet.apa.org/record/2016-01245-001>

¹⁵ <https://www.govinfo.gov/content/pkg/CFR-2006-title29-vol4/xml/CFR-2006-title29-vol4-sec1607-16.xml>

¹⁶ <https://www.govinfo.gov/content/pkg/CFR-2014-title29-vol4/xml/CFR-2014-title29-vol4-part1607.xml>

¹⁷ <https://www.mckinsey.com/industries/private-equity-and-principal-investors/our-insights/the-state-of-diversity-in-us-private-equity>

Summary:

No areas of potential bias or adverse impact were identified in the Crucible assessment. While there were small group mean score differences detected, the differences favor populations that are grossly underrepresented in private equity and thus do not create a concern for adverse impact.

The testing will be regularly reviewed, and the population means reassessed, to ensure potential biases and any adverse impact continue to be minimized. The Crucible strives to account for biases to enable individuals of varying identities to develop in environments that historically have lacked diversity.

APPENDIX:

The following set of tables contain the t-tests (two-tailed) for The PEQ score and Catalyst level comparing female and male respondent group mean scores.

TABLE B1. Female vs. Male - PEQ

Group	Mean	Variance	n	t	p	Cohen's d
Female	33.2	269.9	213	-0.20	0.840	-0.02
Male	33.6	210.8	92			

TABLE B2. Female vs. Male - Execution Engine

Group	Mean	Variance	n	t	p	Cohen's d
Female	30.1	325.9	213	0.43	0.667	0.05
Male	29.2	251.4	92			

TABLE B3. Female vs. Male - Emotional Leverage

Group	Mean	Variance	n	t	p	Cohen's d
Female	30.5	318.8	213	-1.30	0.194	-0.16
Male	33.2	258.6	92			

TABLE B4. Female vs. Male - Leadership

Group	Mean	Variance	n	t	p	Cohen's d
Female	37.3	331.0	213	-0.67	0.50	-0.08
Male	38.5	192.5	92			

TABLE B5. Female vs. Male - Cognitive Horsepower

Group	Mean	Variance	n	t	p	Cohen's d
Female	38.5	731.3	213	-0.10	0.924	-0.01
Male	38.8	658.8	92			

TABLE B6. Female vs. Male - Interpersonal

Group	Mean	Variance	n	t	p	Cohen's d
Female	33.7	465.3	213	2.12	0.036*	0.26
Male	27.9	478.9	92			

* = statistically significant with $p < .05$.

The following set of tables contain the t-tests (two-tailed) for Elements within the Interpersonal Catalyst comparing female and male respondent group mean scores.

TABLE B7. Female vs. Male - Self Esteem

Group	Mean	Variance	n	t	p	Cohen's d
Female	38.5	409.5	213	0.10	0.920	0.01
Male	38.3	312.8	92			

TABLE B8. Female vs. Male - Belief in Others

Group	Mean	Variance	n	t	p	Cohen's d
Female	38.0	777.7	213	1.41	0.160	0.17
Male	33.3	681.1	92			

TABLE B9. Female vs. Male - Empathy

Group	Mean	Variance	n	t	p	Cohen's d
Female	38.4	614.7	213	2.98	0.003*	0.36
Male	29.8	496.2	92			

TABLE B10. Female vs. Male - Confidence

Group	Mean	Variance	n	t	p	Cohen's d
Female	34.9	403.0	213	0.31	0.757	0.04
Male	34.2	281.0	92			

TABLE B11. Female vs. Male - Authenticity

Group	Mean	Variance	n	t	p	Cohen's d
Female	30.6	681.4	213	2.85	0.005*	0.35
Male	21.7	596.4	92			

TABLE B12. Female vs. Male - EQ

Group	Mean	Variance	n	t	p	Cohen's d
Female	34.9	594.7	213	1.61	0.108	0.19
Male	30.5	430.4	92			

TABLE B13. Female vs. Male - Collegiality

Group	Mean	Variance	n	t	p	Cohen's d
Female	33.1	810.4	213	-0.24	0.810	-0.03
Male	34.0	784.9	92			

* = statistically significant with $p < .05$.

The following set of tables contain the t-tests (two-tailed) for Elements within the Interpersonal catalyst comparing Black and White respondent group mean scores.

TABLE B14. Black vs. White - PEQ

Group	Mean	Variance	n	t	p	Cohen's d
Black	34.6	278.0	77	1.18	0.239	0.18
White	31.8	217.0	100			

TABLE B15. Black vs. White - Execution Engine

Group	Mean	Variance	n	t	p	Cohen's d
Black	31.4	323.7	77	0.33	0.739	0.05
White	30.5	287.7	100			

TABLE B16. Black vs. White - Emotional Leverage

Group	Mean	Variance	n	t	p	Cohen's d
Black	33.1	285.6	77	0.20	0.837	0.03
White	33.0	272.9	100			

TABLE B17. Black vs. White - Leadership

Group	Mean	Variance	n	t	p	Cohen's d
Black	39.4	271.0	77	1.43	0.155	0.21
White	36.2	229.6	100			

TABLE B18. Black vs. White - Cognitive Horsepower

Group	Mean	Variance	n	t	p	Cohen's d
Black	35.6	644.9	77	0.71	0.479	0.10
White	32	597.5	100			

TABLE B19. Black vs. White - Interpersonal

Group	Mean	Variance	n	t	p	Cohen's d
Black	32.8	597.5	77	1.99	0.048*	0.30
White	26.1	422.5	100			

* = statistically significant with $p < .05$.

The following set of tables contain the t-tests (two-tailed) for Elements within the Interpersonal Catalyst comparing Black and White respondent group mean scores.

TABLE B20. Black vs. White - Self Esteem

Group	Mean	Variance	n	t	p	Cohen's d
Black	40.7	419.6	77	1.36	0.177	0.20
White	36.8	276.1	100			

TABLE B21. Black vs. White - Belief in Others

Group	Mean	Variance	n	t	p	Cohen's d
Black	34.5	885.2	77	0.46	0.648	0.07
White	32.6	597.5	100			

TABLE B22. Black vs. White - Empathy

Group	Mean	Variance	n	t	p	Cohen's d
Black	39.0	589.7	77	1.18	0.240	0.17
White	34.5	669.0	100			

TABLE B23. Black vs. White - Confidence

Group	Mean	Variance	n	t	p	Cohen's d
Black	35.4	398.9	77	0.81	0.418	0.12
White	33.1	284.8	100			

TABLE B24. Black vs. White - Authenticity

Group	Mean	Variance	n	t	p	Cohen's d
Black	24.6	627.0	77	1.23	0.219	0.19
White	20.0	578.0	100			

TABLE B25. Black vs. White - EQ

Group	Mean	Variance	n	t	p	Cohen's d
Black	36.8	705.1	77	2.65	0.009*	0.41
White	27.0	467.3	100			

TABLE B26. Black vs. White - Collegiality

Group	Mean	Variance	n	t	p	Cohen's d
Black	37.2	934.7	77	1.97	0.050*	0.30
White	28.5	713.1	100			

* = statistically significant with $p < .05$.