OPTO – OBJECTIVITY IN ORGANISATIONAL TEST DESIGN

ABOUT THE GENERAL FOCUS OF OBJECTIVITY IN THE DESIGN PROCESS OF OPTO AND HOW STUDIES ON FAIRNESS CONCLUDE OPTO AS UNBIASED IN RELATION TO ETHNICITY.

OBJECTIVITY IN TESTING

The fundamental nature of using assessment tests, is to create objective and unbiased data, upon which one can rely to make sound, data-driven, and non-discriminating decisions. Therefore, focusing on the test being non-biased towards gender, age, ethnicity etc. becomes of great importance from the very beginning of test development. When it has been successful in the development process, it becomes evident in later studies conducted to validate the test.

Ensuring objectivity and fairness in testing is about being objective, avoiding bias, and not discriminating. A first important step towards this focus must be taken early in the development process of a test resulting in avoiding systematic bias.

If systematic biases are affecting an assessment procedure, not only might the processes be unethical or even illegal, but it may also prevent recognising talent. Using an unbiased test along with a structured approach based on data is a big step towards minimising risks of discrimination that can arise from personal preferences or hidden prejudices.

BIASES OR GROUP DIFFERENCES

Sorting data collected from assessments is likely to show differences between groups. These differences can occur as a matter of chance, they can be representations of actual pre-existing differences between the groups, or they can be caused by deficiencies in the test that systematically favour or disfavour a group (Schmidt, 1988).

Biases are defined as variations in test results which are not representations of real-life differences. Bias can be the results of deficits in the tests, misinterpretation of results, or simply the human factor that plays a misfortunate role in the process.

A test can show difference between two groups on average without it being bias. But if the test overestimates or underestimates one of the groups, the test is biased. Statistically speaking bias arises when there is a factor inherent in a test that systematically influence the test score and prevents accurate and impartial measurement. This can be investigated using various methods and the result of these is the focus of this article.

As mentioned, there are actual differences which represents existing differences between groups. These group differences *should* be found and are evidence that the test *is* measuring correctly.

As such, population variances within personality are likely to exist, indicating that existing differences between groups can occur without it being bias. Because group differences can influence recruitment, it requires professionals to be aware of them and how they can affect their recruitment process.

Big Five and fairness in testing

Various studies have shown group differences within the Big Five Dimensions[1] and their underlying traits (De Young, 2014; Costa, Terracciano & McCrae, 2001; Goldberg, 1998; Soto et al., 2011).

Specific traits of personality, that are linked to certain life outcomes may differ across populations, due to differences in socialisation experiences or other

---

[1] OPTO is a personality test based on the Big Five model or Five Factor of personality (McCrae & Costa, 1999).

master™
VALUE PEOPLE
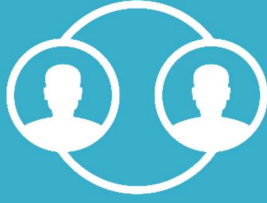
OPTO – OBJECTIVITY IN ORGANISATIONAL TEST DESIGN

ABOUT THE GENERAL FOCUS OF OBJECTIVITY IN THE DESIGN PROCESS OF OPTO AND HOW STUDIES ON FAIRNESS CONCLUDE OPTO AS UNBIASED IN RELATION TO ETHNICITY.

biosocial factors. For example, Agreeableness tend to be somewhat higher among women, older adults, and in some cultures (Soto et al, 2011).

All in all, Soto (2021) concludes though from his meta-analysis[2] on Big Five, that:

> *"... ethnicity only modestly affected the average trait-outcome association and rarely had a meaningful impact on individual associations."*

This underlines the importance of the test documentation including evidence of expected differences across demographical groupings to account for test-specific variances, as is the case with OPTO.

OBJECTIVITY IN OPTO DESIGN

In an extensive review on test biases for minority groups, Sackett et al. (2001) account for possible biases when using tests. They found that while personality tests are generally a lot *less* biased than other types of tests, there are several ways to counter discrimination, areas which are carefully considered in the design of OPTO.

IN OPTO THE FOLLOWING AREAS ARE RELEVANT:
- Minimise reading requirements.
- Reduce verbal complexity of the included items.
- Eliminate stress by not limiting time.
- Ensure that Test Takers are motivated.
- Eliminate any meaningful differential item functioning.

Sackett et al. (2001)

OPTO considers various concerns to ensure test fairness. For instance, OPTO was designed specifically to minimise bias derived from having inadequate language skills by employing short and simple Likert-type statements.

There is no time-limit on OPTO that can potentially induce stress. It can also be assumed that Test Takers are highly motivated in a recruitment scenario. Additionally, motivation is thought into the test experience and flow presenting a motivational text halfway through the questionnaire.

The items[3] of the test are continually monitored, and in cases where items are flagged for potential issues, a review of logical evidence of bias is performed to be able to determine how to proceed with that item.

Updating the design

In 2020 new items were introduced, tested, and implemented in OPTO to strengthen the psychometric properties of the test. The upgrade contained optimisation on item content, introducing new items and retaining only the items with the best psychometric properties. This included removal of a few old items. The upgrade did not introduce any changes to the structure of the test but enhanced the precision of the measurement. The items are afterwards even more closely linked to business, and customers will have an even more precise measurement for selection of candidates.

The total amount of items is only marginally changed, as the upgrade resulted in going from 155 to 154 items, and the Test Taker experience has therefore not been affected by the updated item-pool.

---

[2] A meta-analysis combines the results of multiple scientific studies conducted as different research studies, but addressing the same issue, into one comprehensive statistical analysis.

[3] A test item is the basic unit of interaction with the test taker. In OPTO it is constructed as a statement (e.g., "I like to take it easy"). In other tests, an item can take on other forms such as questions, tasks etc.
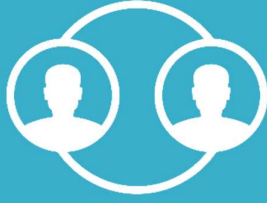
Studies show that applying **high quality objective assessment tools**, such as OPTO, can minimise the risk of poor hiring decisions due to judgement and evaluation bias.

master™
VALUE PEOPLE

LEARN MORE ABOUT OUR SOLUTIONS AT MASTER-HR.COM

OPTO – OBJECTIVITY IN ORGANISATIONAL TEST DESIGN

ABOUT THE GENERAL FOCUS OF OBJECTIVITY IN THE DESIGN PROCESS OF OPTO AND HOW STUDIES ON FAIRNESS CONCLUDE OPTO AS UNBIASED IN RELATION TO ETHNICITY.

A practical feature of OPTO, is that it is possible to leave out Demographic information from the reports. This strengthens the fairness of the test, and potentially removes risk of discriminating, when reading or presenting the results.

STUDIES ON OPTO FAIRNESS

Master International has conducted group-level analyses on OPTO for a variety of demographic variables to bring forth any possible differences between groups and potential biases that could favour one group over the another.

Studies have been conducted to look for bias, and the main conclusion can be summarised as follows:

"The implications for using OPTO Aspects across groups are minor. The group differences are typically less than 1 STEN[4], and they do reflect the population adequately. As a result, the differences in Aspect scores across groups is due to actual group differences and not, for example, a lack of validity (or reliability) on the part of OPTO."

In general, the studies show no evidence of item or test differences. Additionally,

the maximum observed group differences across demographic groups are usually around 1 STEN, making the practical impact on any group-level analysis very small. Thus, there is evidence supporting that OPTO ensures a fair recruitment process across people of different age, educational level, ethnicity, and gender.

In addition to studies the test is designed to be used across cultures. Ensuring fairness has been an integral part of the test design process, supporting the non-discriminating use of OPTO.

ETHNICITY AND TEST DESIGN

To address possible inconsistencies in OPTO assessment responses, any cultural aspect of items has been considered in the design of OPTO. The national translations of the test have been through an extensive adaptation process designed to minimise any cross-cultural bias. Studies have been conducted to ensure there are no bias in

items between languages. All items were originally designed to avoid any highly cultural-dependent content and idiomatic language. Accordingly, OPTO is regarded as unbiased in relation to culture.

Norms

To address any group differences between languages and/or cultures norms are developed and kept up to date.

Country-based norms are created to be able to assess differences and variances within and between countries and national identities, thus, ensuring cultural fairness. For this reason, only, the Test Taker is asked to provide information on their country of origin. Importantly, this information does not influence the individual's result. It is self-reported (Wang and Sue, 2005), and can be left out in the report, as previously mentioned.

The Test Taker's chosen country of origin does not determine the norm used to calculate their OPTO score. It might often be the case that they are identical, but it can be relevant to choose a different norm than the Test Taker's chosen country of origin. The norm group (e.g., US, Polish, or French) is defined as the group of people to whom the Test Taker

[4] "Standard Ten score" showing results on a simple standardized scale from 1 to 10 following the normal distribution.

Studies show that applying high quality objective assessment tools, such as OPTO, can minimise the risk of poor hiring decisions due to judgement and evaluation bias.
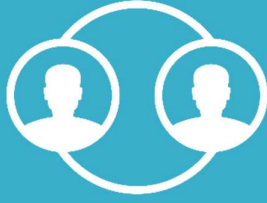
OPTO – OBJECTIVITY IN ORGANISATIONAL TEST DESIGN

ABOUT THE GENERAL FOCUS OF OBJECTIVITY IN THE DESIGN PROCESS OF OPTO AND HOW STUDIES ON FAIRNESS CONCLUDE OPTO AS UNBIASED IN RELATION TO ETHNICITY.

is compared to and is unlinked to the individual demographics of the Test Taker. For example, if hiring a person to be a part of a French team, it can be relevant to use a French norm regardless of the nationality and culture of the candidate.

Asking for respondent's ethnicity

The legislation on asking about ethnicity in a survey-type questionnaire – like OPTO - differs between countries. As a main rule within the EU, sensitive information cannot be included in the test, and ethnicity can in some cases be considered as such. Therefore, it is hard to gather data on ethnicity, and if data exists it is, legally speaking, problematic to use. Consequently, data on differences in scores between ethnic groups are limited and therefore OPTO norms do not feature any weighing for ethnicity.

The studies that have been conducted on test fairness of ethnicity have been based on anonymised data collected by a third party for the sole purpose of conducting these studies.

STUDY ON OPTO ETHNICITY FAIRNESS

An OPTO study conducted in 2017 examining group-level differences across ethnicity in the USA (n=951). The study only found minor group differences, which indicates no impact on the fair use of OPTO across ethnic groups.

The data for the study was collected in October and November 2017 by an external vendor. The sample consists of volunteer panellists. Based on demographic keying, it was possible to compare the sample to the general working population.

In general, the sample is considered representative of the general working population in the US. The population statistics are taken from the official population census estimates (2016) at http://www.census.gov/.

Gender	Population	Sample
Male	49.2 %	47.6 %
Female	50.8 %	52.4 %

The gender distribution of the data sample reasonably resembles the population distribution.

Regarding ethnicity the sample is not fully representative for the ethnic composition of the US population. In the sample White Americans are underrepresented and Asian Americans over-represented. This was sampled intentionally with the purpose of having sufficient data in each category to investigate group differences.

To ensure a balanced sample, analysis of demographic group differences, other than ethnicity, were performed. To assess the occurrence of normed score differences across ethnic groups, the study used an ANOVA design with a Games-Howell Post Hoc test.

The study showed significant differences across ethnicity for Compliance, Innovation and Resilience. These were analysed using the Post Hoc test, looking for significant differences on all OPTO dimensions.

Ethnicity	Population*	Sample
Asian American	5.7 %	12.0 %
Black or African American	13.3 %	16.4 %
Hispanic or Latino American	17.8 %	18.0 %
White American	61.3 %	48.2 %
Americans with Mixed or Multiple Ethnical Groups	2.6 %	3.9 %
*20 to 69 years old		

Studies show that applying **high quality objective assessment tools**, such as OPTO, can minimise the risk of poor hiring decisions due to judgement and evaluation bias.
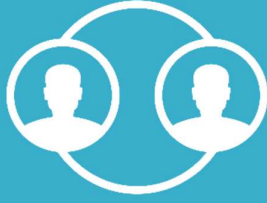
OPTO – OBJECTIVITY IN ORGANISATIONAL TEST DESIGN

ABOUT THE GENERAL FOCUS OF OBJECTIVITY IN THE DESIGN PROCESS OF OPTO AND HOW STUDIES ON FAIRNESS CONCLUDE OPTO AS UNBIASED IN RELATION TO ETHNICITY.

The significant differences were:

- White Americans scored **higher** than Hispanic or Latino Americans (0.51 STEN) on **Compliance**
- Black or African Americans scored **higher** than White Americans (0.48 STEN) on **Innovation**
- Americans with Mixed or Multiple Ethnical Groups scored **higher** than White Americans (0.84 STEN) on **Innovation**
- Black or African Americans scored **higher** than White Americans (0.70 STEN) on **Resilience**
- Asian American scored **higher** than White Americans (0.71 STEN) on **Resilience**

All the significant differences were less than 1 STEN, which indicates that any group differences are unlikely to have a meaningful effect on conclusions made from OPTO results. Furthermore, the differences are not uniform, which is evidence that there are no consistent bias present (e.g., if Whites Americans always scored lower or higher than others).

IMPLICATIONS FOR USE

One method for ensuring fair use of tests in recruitment is to use the same objective criteria for all applicants, and preferably to set the criteria in advance. OPTO takes advantage of this method by allowing the Test User to set criteria for the optimal profile on the Master Test Platform.

Having high scores on OPTO Dimensions or Aspects is not inherently "good" or "bad". It depends on what Test Users are searching for in new candidates. That way, OPTO has no indigenous selection based on scores and thus, no inherent bias. This is key to ensuring test fairness since OPTO has no notion of which profiles are good or bad.

The definition of what is a desired trait, is left completely up to the definition of the job situation. And as such, any bias introduced is done solely by the Test Users themselves. The Master Test Platform also offers the option to create Criteria automatically based on a selected group (e.g., high performers). This gives the opportunity to step away from personal judgements of the optimal profile, towards implementing a more un-biased data-driven approach and thereby avoiding discrimination in recruiting.

CONCLUSION

OPTO can be applied fairly across ethnical groups as a non-discriminating tool for selection and screening. Evidence suggests that the differences found are small enough to claim that OPTO scores equivalently across ethnical groups, and the evidence does not suggest that there are special considerations to be made in relation to applying the norms when using the test for a diverse candidate pool.

As studies on ethnicity do not find an effect of ethnicity on scores in OPTO, it is regarded as irrelevant to ask for the Test Takers' ethnicity in an assessment situation.

Users of OPTO can rest assured that several measures have been taken on behalf of Master International to avoid bias and ensure fairness:

- OPTO is based on Big Five theory, where ethnicity rarely has a meaningful impact on individual results (Soto, 2021).
- OPTO is designed with a focus on avoiding bias following scientific recommendations (Sackett et al., 2001).
- Studies on OPTO fairness have been conducted showing that the implications for using OPTO Aspects across groups are minor, and OPTO is regarded as unbiased in relation to ethnicity (cf. OPTO studies on ethnicity fairness, 2017).
- OPTO is continuously developed to ensure high quality, assuring test and results continue to be unbiased and supporting non-discriminating use (cf. latest OPTO update, 2020).

Studies show that applying **high quality objective assessment tools**, such as OPTO, can minimise the risk of poor hiring decisions due to judgement and evaluation bias.

master
VALUE PEOPLE

LEARN MORE ABOUT OUR SOLUTIONS AT MASTER-HR.COM
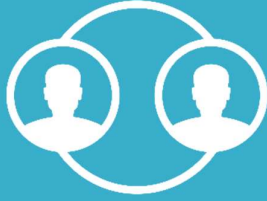
OPTO – OBJECTIVITY IN ORGANISATIONAL TEST DESIGN

ABOUT THE GENERAL FOCUS OF OBJECTIVITY IN THE DESIGN PROCESS OF OPTO AND HOW STUDIES ON FAIRNESS CONCLUDE OPTO AS UNBIASED IN RELATION TO ETHNICITY.

REFERENCES

Aguinis, H., Culpepper, S. A. & Pierce, C. A. (2010) Revival of Test Bias Research in Preemployment Testing. Journal of Applied Psychology. Volume 95(4) https://doi.org/10.1037/a0018714.

Census estimates on official US population. http://www.census.gov/

Costa, P. T., Terracciano, A., & McCrae, R. R. (2001) Gender differences in personality traits across cultures: Robust and surprising findings. Journal of Personality and Social Psychology. Volume 81(2): 322–331. https://doi.org/10.1037/0022-3514.81.2.322

DeYoung, C. (2014) Openness/Intellect: A dimension of personality reflecting cognitive exploration. APA Handbook of Personality and Social Psychology: Personality Processes and Individual Differences. Volume 4: 369–399. http://doi.org/dx.doi.org/10.1037/14343-017

Goldberg, L. R., Sweeney, D., Merenda, P. F., & Hughes, J. E. (1998) Demographic variables and personality: The effects of gender, age, education, and ethnic/racial status on self-descriptions of personality attributes. Personality and Individual Differences. Volume 24(3): 393–403. https://doi.org/10.1016/S0191-8869(97)00110-4

McCrae, R. R., & Costa, P. T., Jr. (1999) A Five-Factor theory of personality. In Pervin, L. A. & John, O. P. (Eds.) Handbook of personality: Theory and research, 2nd edition, pp. 139-153. Guilford.

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001) High-stakes Testing in Employment, Credentialing, and Higher Education: Prospects in a Post-Affirmative-Action World. American Psychologist, APA.

Schmidt, F. L (1988) The Problem of Group Differences in Ability Test Scores in Employment Selection. Journal of Vocational Behavior. Volume 33: 272-292

Soto, C. J. (2021) Do link between personality and life outcomes generalize? Testing the robustness of trait-outcome associations across gender, age, ethnicity, and analytic approaches. Social Psychology and Personality Science. Volume 12: 118-130

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011) Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. Journal of Personality and Social Psychology. Volume 100(2): 330–348. https://doi.org/10.1037/a0021717

Wang, V. O. & Sue, S. (2005) In the Eye of the Storm – Race and Genomics in Research and Practice. American Psychologist. Volume 60(1): 37-45 https://doi.org/10.1037/0003-066X.60.1.37

Studies show that applying high quality objective assessment tools, such as OPTO, can minimise the risk of poor hiring decisions due to judgement and evaluation bias.

master™
VALUE PEOPLE

LEARN MORE ABOUT OUR SOLUTIONS AT MASTER-HR.COM