

Translated Languages: Answers to Frequent Questions about the Denison Organizational Culture Survey

July, 2009



Bringing Organizational Culture and Leadership to the **Bottom Line**.

Globalization has created a seemingly difficult dilemma for conducting organizational surveys: can survey items be created that will mean the same to a technology worker in Seattle as they do to a pharmacist in Berlin? In other words, will those items mean the same across one of the most fundamental differences: language? The Denison Organizational Culture Survey (DOCS) has been translated into 38 languages, and this White Paper offers insight into the functioning of the DOCS across these languages.

While many organizational surveys are translated into multiple languages, there is often little evidence to support that the survey is equivalent across those translations. Because organizations count on survey results for everything from informing organization-wide change efforts and policy decisions to integrating organizations after a merger, it is critical that the items from these surveys mean the same thing regardless of whether employees took the survey in English or Japanese. To this end, Denison Consulting worked with researchers at Bowling Green State University in Bowling Green, OH, USA to investigate the translations of the DOCS and determine their validity.

Following are some questions that users of the Denison Culture Survey may have regarding our translations process.

1. How was the Denison Culture Survey translated into different languages?

The US English scale was translated by individuals fluent in both English and the translated language. We used a translation/back translation process where the survey was first translated from US English to the target language, and was then re-translated back into US English by a different person fluent in that language. This ensured that the items did not lose their meaning. Any questionable translations were discussed and resolved.

2. Have the translations been localized?

As often as possible, yes. The translations have been reviewed by native speakers, which has been extremely helpful to ensuring that the meaning of each item is understood from the perspective of the survey-taker. When needed, revisions have been made based upon the evaluations of native clients.

Bringing Organizational Culture and Leadership to the **Bottom Line**.

3. How do you determine if the translations are good?

There are a number of different statistical techniques that can be employed to determine the quality of a translated survey. Choosing which technique to use is primarily determined by the amount of available data. A technique called Item Response Theory (IRT) is the “gold standard” for evaluating translation equivalence but can only be used with samples of 300 respondents or more. The specific IRT analysis used by the researchers from Bowling Green State University is called Differential Item Functioning (DIF). DIF analysis examines the difference between two people who have the same view of the organization’s culture but who took the survey in two different languages: one in US English, versus another translated language. A second technique that can be used to supplement the IRT analysis is Confirmatory Factor Analysis (CFA) which determines if the Denison Culture items have the same meaning in the translated language as it does in US English.

4. What were the results of the “Gold-Standard” analyses?

For the work conducted by Bowling Green State University researchers, only 14 languages had more than 300 respondents with which advanced IRT and CFA analyses could be used to compare these translations to the US English form. In conducting these analyses we were looking to determine if we have an apple-to-apple comparison. Basically, if we give the Denison Organizational Culture Survey in French do we get a French employee’s beliefs and assumptions of their organization’s culture or do we get something else?

First we used IRT to make comparisons between the US English version and the 14 other languages. Approximately 840 comparisons were made at the item level (60 items x 14 translations) and 168 comparisons were made at the index level (12 indexes x 14 translations). The results of the IRT analyses indicate that the vast majority of items and indexes are comparable to the US English version – we gave apples and we got apples back in return. Overall, 88% of the items and 166 index translations were statistically equivalent. This represents an impressive degree of equivalence given the number and

Bringing Organizational Culture and Leadership to the **Bottom Line**.

complexity of items (60 per language), as well as the number of different languages (14) being compared. At the index level, where 168 comparisons were made, the two exceptions were the French version of *Empowerment* and the German version of *Customer Focus*. Although these results were surprising, the severity of this finding is similar to finding a brown spot on an apple – it does not mean that the whole apple is bad but that spot should probably be cut out. Luckily with translations we can “fix” the brown spots on our surveys!

We also conducted 14 CFA tests to determine if the Denison Culture items have the same structure in the translated language as it does in US English. Results indicated that the same indexes and culture traits were found across the translated languages. The one exception was the Turkish version of the culture survey which indicated that the responses of Turkish participants were not as reliable as those of the US English version. Returning to our apple metaphor this would be similar to comparing a Granny Smith apple to a Golden Delicious apple – they are both apples but slightly different. Given that the IRT analyses for the Turkish translation did not reveal any problems we feel that this is a relatively minor result.

5. Why couldn't you evaluate all of the translations using these advanced statistics?

One challenge to statistically determining translation quality is having enough data collected across languages that are also collected from the same organization. These types of analyses compare the responses of people taking the translated survey to the responses of people taking the survey in US English. This allows us to be certain that the differences we detect are due to item translation and not a comparison of culture in different organizations. This does, however, limit the amount of data that is available to work with. There were 14 languages for which the sample sizes were not large enough to conduct the more advanced analyses previously discussed. Therefore, we used basic statistics to determine if there were any languages or items that might be

Bringing Organizational Culture and Leadership to the **Bottom Line**.

underperforming and would merit further exploration. In these cases the average sample size for these languages ranged from 33 to 204 people.

Glossary of Statistics Used in Evaluating Denison Culture Items

Term	Definition
Item Response Theory	a framework for evaluating how well individual questions on surveys work
Confirmatory Factor Analysis	a framework for evaluating the structural model of a set of items
Reliability	the “consistency” of a set of items that indicates whether the same result would be achieved each time a survey is administered
Item-Total Correlation	the correlation between an item and the index that it is associated with
Differential Item Functioning	examines the difference between two people who have the same view of the organization’s culture but who took the survey in two different languages

6. What did the basic analyses indicate about the translations?

The two types of basic analyses we conducted were reliability and item-total correlations. First, the reliability of each Index was calculated for each language and compared to the reliability score for that Index in the US English sample. Reliability is often-times referred to as the “consistency” of a set of items and is an indication of whether the same result would be achieved each time a survey is administered. The more reliable a set of items are (the closer the value is to 1.0) the better although anything around .70 or higher is considered pretty good from a statistical point of view. The reliability analyses ranged from .69 to .87 for the 14 languages and from .68 to .85 for the US English sample. Only two indexes had really low reliability scores: the *Agreement* index in the Vietnamese translation (.28) and the *Creating Change* index in

Bringing Organizational Culture and Leadership to the **Bottom Line**.

the Indonesian translation (.32). All of the other indexes had reliability scores in the acceptable range.

Second, item-total correlations were calculated and scanned for inconsistencies. An item-total correlation is the correlation between an item and the index that it is associated with. An item-total correlation functions the same as a correlation (with a range of -1.0 to +1.0) with scores being closer to 1.0 preferred. Approximately 840 item-total correlations were calculated (60 for each language x 14 languages). Only 2 item-total correlations were lower than what is the recommended level in the scientific literature. One item was in the Vietnamese translation and one item was in the Indonesian translation. There were only 33 people who completed the Indonesian translation and 43 people who completed the Vietnamese translation. Therefore, these findings are not conclusive and with a larger sample size these results might be different.

Languages Evaluated by Bowling Green State University Researchers

Languages Evaluated using Advanced Statistics	Languages Evaluated using Basic Statistics
Chinese Simplified	Croatian
Chinese Traditional	Czech
Dutch	Danish
UK English	Finnish
French	Flemish
German	French Canadian
Italian	Hungarian
Japanese	Indonesian
Korean	Norwegian
Portuguese	Polish
Portuguese-Brazil	Russian
Spanish-Latin America	Swedish
Spanish-Non-Latin America	Thai
Turkish	Vietnamese

Bringing Organizational Culture and Leadership to the **Bottom Line**.

7. Since you did all of these analyses in one organization could it be the culture of the organization that produced these results?

Because we conducted these analyses using data from one organization we cannot rule out the suspicion that these results could be due to the culture of that particular organization. To investigate this question further, similar Advanced Statistics were conducted that compared our US English sample to two additional organizations who had taken the survey in US English: one based in the United States and one based in Europe. When we compared our first US English sample with a second sample we found three items that functioned differently in the other organization compared to our US English sample (Remember: apple-to-apple comparisons). When we compared our US English Sample to the second sample, a European-based organization that took the survey in US English, we also found three items that did not perform the same way – but they were not the same three items. Fifty-seven of the sixty items were found to behave the same across languages, which is an extremely high rate of convergence. Also, because the items that were underperforming when compared to our US English sample were not the same between the two comparisons we can conclude that the results found by the Bowling Green State University researchers are not due to the culture of the first organization.

8. So, if I was to explain these results to my grandmother, in a nutshell, what should I tell her?

Whether your grandmother goes by Mema, Grandmama, Oma, Uma, Imo, Nonna, Grand-Mere, or Abuela, you should tell her that the Denison translations are pretty good! Considering all of the analyses together:

- The United Kingdom English version was wholly comparable at both the item and index levels to the US English version
- Eastern/Asian language translations had more non-comparable items
- Similar languages, such as Portuguese and Portuguese-Brazil or Chinese Traditional and Chinese Simplified, behaved in the same way

Bringing Organizational Culture and Leadership to the **Bottom Line**.

9. What are Denison's plans in the future?

The analyses of the translated languages of the Denison Organizational Culture Survey were very promising and indicate that the meaning of the Denison Culture items translate well into other languages, however, there are always opportunities to improve on our translations. The language translations with small sample sizes will be continually evaluated by the Denison Research & Development team and we plan to work with our client organizations to localize the Denison Culture Survey to different languages.

10. What if I want to know more about these analyses?

If you love – and we mean LOVE – statistics you are welcome to read through a technical report that describes, in full detail, the work that was done by Bowling Green State University to evaluate the Culture Survey translations. Please contact a Denison representative to schedule a meeting with a Denison Research Consultant.

Referenced Technical Reports

- 1. An Examination of the Practical Implications of Differential Functioning in the Denison Organizational Culture Survey.* Carter, N.T. (2008). Bowling Green, OH: Bowling Green State University, Department of Psychology.
- 2. Measurement equivalence of the Denison Organizational Consulting Services Survey across language adaptations.* Carter, N.T., Diab, D., Lin, B.C., Pui, S., Zickar, M.J. (2007). Bowling Green, OH: Bowling Green State University, Department of Psychology, and the Institute for Psychological Research Application.