Calibrating Bilingual Q Samples

Steven R. Brown
Kent State University

Ursula Feist
Institut für angewandte Sozialwissenschaft

ABSTRACT: In cross-cultural studies, the comparability of Q samples in different languages becomes problematic, and in this study procedures are introduced for enhancing confidence that parallel Q sorts written in two languages provide roughly equivalent results. 11 English-speaking German nationals performed two Q sorts containing the same statements, one version in German and a parallel version in English, and correlations and factor analysis indicated virtually no differences in outcome. Discussion focuses on the cross-cultural comparability of factors in R and Q methodology.

From Potsdam to Prague, Phnom Penh to Peking, and St. Petersburg to Pyongyang, the increasing openness in the East is multiplying opportunities for collaboration on comparative studies of cross-cultural experience, and the pace is apt to quicken as electronic networks and other cultural tunnels

Authors' addresses: *Brown,* Department of Political Science, Kent State University, Kent, OH 44242-0001; *Feist,* infas GmbH, Büro Hamburg, Gasstr. 18, Haus 4, 2000 Hamburg 50, Germany.

This study was supported by grant SOC74-21584 from the National Science Foundation. The authors are grateful for the critical comments of reviewers.

continue to proliferate. In many instances, comparative studies of this kind require parallel versions of measuring instruments which have been validated for their respective contexts; in the case of Q technique, however, such standardization is forgone, even in monolingual situations. As Stephenson (1961) once said in this regard:

> ... there is no question at all of collecting any "standard" Q-sample; nor is it assumed that the statements mean the same to everyone; nor is any question of validity or reliability "in general" at issue. The concern is with concrete interbehavior.... (p. 21)

The purpose of "equivalent" Q samples in two languages is not to assure standardized meaning, therefore, as in scaling theory, but to provide the basis for some degree of assurance that whatever point of view a person has expressed in a French, German, Hindi, Swahili or other-language Q sort is approximately the same as the view the person would have expressed were there no language barrier. In other words, if person X agrees with statement *a* more than with *b*, are there demonstrable grounds for believing that the same relationship would prevail were everything the same except for the language of the Q sort? This is an especially important consideration for the investigator, who at some point must interpret the results provided by subjects from a different culture, and who may understandably wonder if the ordering of the statements as he or she is observing them (say, in English) might actually be different due to purely linguistic differences between English and the language in which the subject is performing the Q sort. Note that the problem at hand says nothing about whether the Dutch characteristically prefer *a* more than *b* whereas for Canadians the order is reversed, nor about whether *a* means something different in Quito than in Dar-es-Salaam: these matters of "concrete interbehavior" remain, and in fact constitute an important part of what we wish to discover.

An Example: Political Climate

A partially worked example may help clarify what is at issue. In this illustration, 11 English-speaking German nationals provided their viewpoints twice -- once using the original English version and once the German translation of a Q sort comprised of opinions on various social and political matters. More specifically, the Q sample (N = 48) was designed to examine the climate of opinion and was balanced for *values* of welfare and deference, *symbols* of demand, identification, and expectation, and *myths* of ideology and utopia (Lasswell, 1965). The level of generality at which climate was examined can be inferred from representative statements:

> America has reached a very high level of prosperity, health, and education, and we have good reason to be proud of these achievements. (welfare, identification, ideology)
>
> At school, if not before, the American child is stripped of his imagination, his creativity, and his dreams in order to fit him to be a productive unit in a mass technological society. (deference, demand, utopia)
>
> With individualism apparently on the way out, the experience of collectivism may become more attractive to people searching for a way to cope with the stresses of life. (welfare, expectation, utopia)

The statements originally appeared in English. Once the Q sample was selected and edited for administration to a U.S. audience, it was translated (under the supervision of the second author) for simultaneous administration in Germany, with all references to America naturally changed to Germany. The translation might be said to have been closest to the "text-linguistic model" in that differences between the source and target versions were based on pragmatic considerations operating beyond the sentence, or on what are referred to as "suprasentential *textual* factors" (Neubert & Shreve, 1992, p. 23). The two versions of the Q sample are available upon request.

The German Q sort was administered first to respondents 1-5, followed by administration of the English version; the order was reversed for respondents 6-11, thereby controlling for presentation effects. Completion of each Q sort took approximately 30-45 minutes; anywhere between a few hours to three days separated the two sortings.

Table 1
German-English Correlations

	1	2	3	4	5	6	7	8	9	10	11
	German correlations										
1	(96)	60	40	34	29	40	53	31	67	52	55
2	53	(94)	45	32	41	32	50	31	72	44	35
3	42	46	(88)	40	31	36	32	37	46	41	43
4	42	40	32	(90)	13	32	49	10	44	56	41
5	32	36	32	14	(93)	03	55	40	41	32	18
6	50	31	35	42	13	(94)	18	15	39	35	27
7	40	23	31	23	43	23	(70)	41	58	65	50
8	37	37	40	19	46	09	52	(77)	42	31	14
9	66	53	41	36	35	54	46	43	(77)	44	43
10	49	52	43	62	47	45	44	36	47	(90)	54
11	51	40	44	46	08	49	29	18	37	49	(88)
	English correlations										

(decimals to 2 places omitted)

The upper off-diagonal of Table 1 shows the correlations from the 11×11 matrix based on the German version of the Q sample, and the lower off-diagonal the correlations based on the English version. The average correlation from both matrices is identical (M = 0.39), which is an initial if inconclusive hint that similar processes are operating in both cases; moreover, none of the *r*s between any two subjects using the German Q sample is significantly different (even at the $\alpha = .10$ level) from the *r* for the same subject pair using the English Q sample. Of the five *r* differences that exceed one standard error, four are associated with subject no. 7.

Entries in the principal diagonal (in parentheses) are the English-German *rs* for each person, and range from 0.96 to 0.70 with a mean of M = 0.87, which is surprisingly high since reliabilities even under constant conditions are usually around 0.80 (Frank, 1956; Steller & Meurer, 1974). These figures document higher within- than between-subject correlations -- i.e., each person correlates more with him- or herself than with anyone else -- as would be the case (due to specificities) were the two forms of the Q sample evoking similar operant responses.

The average self correlations for respondents 1-5, who took the German version first, are somewhat higher than for respondents 6-11 (0.93 vs. 0.85). This difference is not significant ($t(9) = 2.25$, $p > .05$); however, it suggests that taking the Q sort in one's own language first may help persons think about the issues more clearly and better prepare them to be more consistent when they retake the Q sort in another language.

Table 2
Factor Solution for Table 1

	German		English		h^2	
	GA	GB	EA	EB	G	E
1	62	44	64	41	58	59
2	45	60	51	43	57	45
3	50	35	45	41	37	38
4	67	11	69	13	47	49
5	05	72	09	67	52	46
6	53	06	70	07	29	50
7	48	61	24	62	60	44
8	12	58	11	72	35	53
9	54	61	56	50	66	56
10	68	34	62	44	58	58
11	67	19	69	12	48	49

Table 2 shows the two-factor solutions (principal axis, varimax rotation) for each of the Q-sort performances, Ger-

man and English, and the results are virtually identical: Each solution accounts for 49.6% of total variance, and this outcome supports the conclusion that the two Q samples are performing the same function.

Moreover, a case-by-case assessment of the factor loadings reveals no significant differences. The largest discrepancy is for subject no. 6, who produced loadings of 0.53 and 0.70, respectively, on the first factors of the German and English versions of the Q sample, but even this difference falls far short of significance ($z = 1.66$) (Expositor, 1992). Subject no. 6 also produced the greatest discrepancy in communalities (0.29 vs. 0.50), and whereas this difference reached significant levels ($z = 2.15$, $p < .05$), it was the only one to do so.

The Q-sort statements are themselves subject to standard error estimates since, in this instance, they can take on different values within the factor arrays associated with the two different Q samples. Utilizing $\alpha = .01$, not a single statement within the first German factor received a score that was significantly different from the score received by the same statement in the first English factor, and only one statement received significantly different scores between the German and English versions in the second factors:

> The rich and the big corporations and labor unions have too much power and privilege. (+ 2)
>
> Die reichen und großen Unternehmen der Wirtschaft und die Gewerkschaften haben zuviel Macht und genießen zu viele Vorrechte. (-2)

It is not wholly clear why this statement, when expressed in English, should evoke a more positive response from factor B subjects than when it is expressed in German -- other than the fact that in the German version the "economy" (Wirtschaft) is explicitly mentioned, which may have introduced a conceptual wrinkle absent in the English version. Even if this discrepancy were to prove replicable, however, and not attributable to error, it would still stand as a mere trifle when compared to the robustness of the two main factors; conse-

quently, the overall interpretations of the two factors, which would be based on the patterns among the factor scores in their entirety, would remain largely comparable. Nonetheless, the above result underscores the importance of the quality of translations, and warns of the possible reactions that can be introduced by small differences in wording.

Table 3
Second-Order Analysis

	Correlations				Factors		
	GA	GB	EA	EB	A	B	h^2
GA	--	20	94	20	96	11	93
GB	00	--	17	87	09	93	87
EA	00	-01	--	19	96	09	94
EB	-01	00	01	--	11	93	87
	residual *rs*						

Finally, the similarities between the respective factors are clearly illustrated when the two sets of factor arrays are themselves intercorrelated and factor analyzed, as shown in Table 3. These two second-order factors account for 90.3% of total variance, and the residuals indicate that among these four factor arrays, these two factors alone are at issue.

In combination, therefore, the various results above fortify confidence that the German translation of the Q sample is sufficiently comparable to the English version so that a rough equivalence could be claimed (pending contrary evidence) for any cross-cultural factor similarities that might emerge. Were German and U.S. nationals to emerge on a common factor, for instance, we would be more comfortable in attributing this to a commonality in outlook; likewise, factor differences would be explained more easily as arising from attitudinal differences than from lack of parallelism in translation. Nothing can be vouchsafed, of course: everything depends on the logic that evolves within the context of a particular study. All that can be claimed is that application of procedures such

as those outlined above can reduce uncertainties concerning sources of similarity and dissimilarity in Q-sort responses. In studies of this kind, post-sorting interviews loom especially large for revealing whatever subtleties in meaning and interpretation might exist.

Partial Results

As it turns out, the administration of this Q sample to individuals in both the U.S. (n = 40) and what was then West Germany (n = 50) produced four factors that were virtually identical in both settings. Some of the substantive details have been reported previously (Brown & Feist, 1977), but an initial sense of the factors can be gleaned from the following distinguishing statements (scores for German factors I to IV, respectively):

+5	-3	+1	-2	(6) What may at first appear to be a decline in morals may actually reflect changing traditions. As we become better educated, we become less bound up in tradition and many of our old values change.
-1	+4	-1	-5	(4) When it comes to choosing someone for a responsible position, I prefer a person who is well established and who has achieved a high level of respect from those who must follow him.
+4	+5	-5	+2	(12) I consider myself in alliance with other humanists who share my progressive hopes for much needed changes and a better life for us all.
-2	-2	0	+3	(9) I am genuinely apprehensive that we are headed for another war. It seems inevitable given the deteriorating economic situation, growing unemployment, rising crime, and the spectre of widespread famine.

The subjects comprising the first factor (the *Utopians,* as they were called) were what, in mid-1970s America, were referred

to as countercultural -- mainly youthful, antiwar, antiauthority -- and this factor documents their existence in West Germany at this time as well. The *Ideological* (factor II), on the other hand, retained respect for authority and provided a bulwark against the rising tide. The *Aspiring* were failing to identify with their fellow citizens, as statement 12 indicates, and elsewhere displayed an attitude of cynicism and the kind of self-centered opportunism that to some extent came to characterize the subsequent decade. Finally, the *Apprehensive* (statement 9) were alarmed by current social and technological trends, and, while liberal in their policy preferences, were conservative in terms of their moralism and religious sentiment.

As was asserted previously, the German translation of the Q sample was not designed to provide standardization of meaning, but as a medium for recording relative salience at the level of individuals and factor types. Whether factors are comparable cross-culturally then becomes an a posteriori empirical matter rather than an a priori matter of establishing equivalent meaning. In this particular study, the factors found in the U.S. were also found in Germany, but there was nothing guaranteeing this in advance, nor does this in any way speak to the adequacy of the Q sample translation. For even if perfect cross-national agreement in meaning on each and every statement had been guaranteed before a single Q sort had been solicited, this by itself would have had little if any effect on the number or comparability of the factors which emerged since the factors, like the Q sorts which they subsume, are based not on shared meaning alone, but also on shared *saliences* -- i.e., that statement *a* (whatever its meaning to the person) is judged more salient, or important, than *b* (whatever its meaning).

Conclusion

We note in closing that the issue of factor comparability has been particularly thorny in R methodology, due to the concern with factor invariance from site to site:

> Regardless of the "meaning" a set of dimensions may seem to have, and regardless of the statistical clarity with which a given set of factors may appear to be defined in any single investigation, they can only form a generally useful descriptive system if the basic conditions of factor invariance are adequately met. From one study to another within the domain under consideration, the same set of factors must dependably emerge. If this condition does not prevail, and if the characteristics examined are not chosen on carefully formulated theoretical grounds, descriptions can be framed as well by any arbitrarily chosen set of variables, combined by any sort of whimsy whatsoever. (Peterson & Migliorino, 1967, pp. 216-217)

In R method, of course, there is the additional problem that scales have the same meaning cross-culturally, i.e., meaning must be generally invariant from place to place. This problem is obviated in Q since all that is required is that subjects of the same factor type *behave* in the same way, i.e., that they sort the statements (whatever meaning may be attached to them) in roughly the same order. The cross-cultural patterning of statements is the basic phenomenon of interest, rather than the intrinsic meaning of items. Whether German and American factor mates are indeed similar is then an empirical rather than a definitional matter. Ultimately, however, whatever different understandings may be reached with regard to Q statements remains a matter for more thorough scrutiny.

References

Brown, S.R. & U. Feist (1977, June) *The cross-national study of political climate: Conceptual and operational foundations.* Paper presented at the meeting of the International Communication Association (Political Communication Division), Berlin.

Expositor (1992) A note on measuring changes in Q factor loadings. *Operant Subjectivity, 15,* 56-61.

Frank, G.H. (1956) Note on the reliability of Q-sort data. *Psychological Reports, 2,* 182.

Lasswell, H.D. (1965) The climate of international action. In H.C. Kelman (Ed.), *International behavior* (pp. 339-353). New York: Holt, Rinehart & Winston.

Neubert, A. & G.M. Shreve (1992) *Translation as text.* Kent, OH: Kent State University Press.

Peterson, D.R., & G. Migliorino (1967) The uses and limitations of factor analysis in cross-cultural research on socialization. *International Journal of Psychology, 2,* 215-220.

Steller, M. & K. Meurer (1974) Zur reliabilität eines Q-sort zur veränderungsmessung. *Psychologische Beiträge, 16,* 618-624.

Stephenson, W. (1961) Scientific creed--1961: The centrality of self. *Psychological Record, 11,* 18-25.