THE USE OF VERBAL LABELS IN NOISE ANNOYANCE SCALES
THEORETICAL DELIBERATIONS AND EMPIRICAL FINDINGS

B. Rohrmann

Dept. of Psychology, University of Melbourne, Australia
E-mail: rohrmann@unimelb.edu.au
WebSite: http://www.psych.unimelb.edu.au/staff/rohrmann.html

0 Summary

Firstly, criteria for designing response scales and selecting verbal scale point labels for annoyance scales are discussed. Then findings from a psychometric study are presented which can be utilized to refine rating scale construction.

1 The issue: Quantifying annoyance

In social-scientific research on the effects of noise, annoyance due to noise exposure is one of the core topics. Virtually all studies, whether lab experiments or population surveys, employ scales to measure this construct.

Two main types of annoyance scales are: direct scaling, i.e., the respondent expresses her/his degree of annoyance (or disturbance, apprehension, irritation etc) on an accordingly anchored and labeled scale; and scaling via statements, i.e., the respondent assesses to which extent s/he experiences particular described noise effects (usually as intensity or frequency). In psychometric terms, the measurement approach is predominantly category scaling (based on differences), but magnitude scaling (based on ratios) is utilized as well. As considerable context effects occur, some researchers have tried to incorporate well-defined reference levels of noise effects and/or to calibrate annoyance judgments. Quantifying annoyance is indispensable for two tasks: to identify individual levels of noise impacts, and to operationalize the noise problem for populations (e.g., "% highly annoyed" in the vicinity of an airport), and thus requires proper scaling methodology. If cross-national comparability is sought (cf. Rohrmann 1985), the issue becomes very complex.

2 Designing response scales

A annoyance response scale should fullfill psychometric standards of measurement quality as well as practicality criteria, such as comprehensibility for respondents and ease of use.

Within category scaling, verbal labeling of rating scales has become the

dominant approach to enhancing usability. The labels are used as "qualifiers, "multipliers", "quantifiers" for particular levels of the issue to be judged (see, e.g., Likert 1932, Cliff 1959, Moxley & Sanford 1993), either for the scale endpoints (e.g., "not-at-all"..."extremely" or "never ... always" for a 0..10 scale); or for each single scale point (e.g., "never/seldom/sometimes/often/always" or "not/ slightly/fairly/quite/very" annoyed). Verbal labeling provides major advantages, such as ease-of-explanation and familiarity (in fact most people prefer verbal responses); it also facilitates to capture normative judgments. The main disadvantage is inferior measurement quality; also, cultural factors might confound the data. Furthermore, cross-national comparability is difficult.

It is therefore essential to design verbalized scales very carefully if equidistant and unambiguous instruments are to be achieved - if possible based on psychometric data for scale labels. A few authors provide such information (e.g. Jones & Thurstone 1969, Hammerton 1986). Within noise research, Rohrmann scaled 100 (German) expressions for a project on aircraft noise effects in 1966 and replicated the study in 1976. The results were utilized to construct (quasi-) intervall scales (some of them seem still to be in use).

3 A study on verbal scale point labels

In 1997-8, a psychometric study was conducted to clarify the measurement features of (English) verbal scale point labels relevant for questionnaire construction and to develop methodologically sound response scales which are useful for both basic and applied research (Project VQS, Rohrmann 1998). The project deals with rating scale construction in general; however, noise annoyance is the main substantive topic utilized in a series of 6 sub-studies.The research design is as follows:

> *Qualifier dimensions:* Five, i.e.:Frequency *(*e.g., never, often); I*ntensity (*e.g., somewhat, very); Probability *(*e.g., unlikely, possibly); Quality *(e.g., bad, good);* Response to statements *(*e.g., disagree, true for me).

> *Number of considered items:* 12+22+16+22+22 = 94 verbal scale point labels (VSPLs), words or expressions, were tested.

> *Scaling tasks:* CAT: Categorial, VSPLs to be placed on a 11-point "equal appearing interval scale" (sensu Thurstone 1929). PREFL: Choosing preferred scale point labels for a 1-2-3-4-5 scale. FAM: Rating familiarity of VSPLs. MAG-N and MAG-L: Magnitude estimations (cf. e.g. Wegener 1983) of VSPLs collected in two modalities, (1) numbers and (2) lines.

> *Contexts for VSPLs:* Noise (e.g.: "I am *<intensity-qualifier>* annoyed by noise; "traffic noise *<frequency--qualifier>* disturbs me; etc); Job satisfaction; and VSPLs presented 'pure' *without* context.

> *Samples:* Students (N=4x30); general population (N=30+30).

A cross-national extension with data collections in Germany (scaling of homologous VSPL's in German language) is currently in preparation.

Selected results - only for noise and the 22 "intensity" VSPLs; only preliminary magnitude scale scores - are summarized in the table below. (Further analyses of the magnitude scaling data, including log-linear transformations and cross-modality matching, are still under way).

SCALING VERBAL QUALIFIERS: SELECTED RESULTS FOR "INTENSITY"

Scaling task	CATEGORIAL (0…10 scale)				MAGNIT-UDE <#>		PREFERED LABEL (% respondents)					FAMILIA-RITY	
Context:	noise		all		all		for annoyance scale level					noise	
	M	sd	M	sd	M	sd	1	2	3	4	5	M	sd
Verbal label													
a little	2.5	1.3	2.5	1.4	10	17		13				7.1	2.7
average	4.7	1.0	4.8	0.9					28			8.8	1.0
completely	9.8	0.6	9.7	0.8	81	161					40	8.5	1.6
considerably	7.5	1.2	7.6	1.1	57	129				21		6.3	1.7
extremely	9.6	0.6	9.6	0.8	76	145					47	8.3	1.4
fairly	5.1	1.3	5.4	1.4	46	113						6.4	1.8
fully	9.2	1.2	9.3	1.3	78	161							
hardly	1.6	1.4	1.7	1.2	9	17		18				7.1	1.8
highly	8.6	0.7	8.6	0.9	68	130						7.4	2.1
mainly	6.4	1.1	6.1	1.4	58	129				18		7.4	1.6
medium	4.8	0.8	4.9	0.8					25			7.3	2.3
moderately	4.9	1.3	5.1	1.1	43	112			37			6.5	2.0
not	0.4	0.5	0.5	0.9	2	3	17					9.4	1.0
not at all	0.1	0.4	0.2	1.0	1	0	70					9.1	1.5
partly	3.5	1.4	3.8	1.4	21	49		14				7.0	1.8
quite	6.1	1.5	5.9	1.5	38	81						6.5	2.4
quite a bit	6.4	1.7	6.5	1.6	45	97							
rather	5.9	1.7	5.8	1.6	46	113						5.7	2.3
slightly	2.5	1.4	2.3	1.5	12	17		27				6.9	1.8
somewhat	4.3	1.7	4.5	1.7	27	49						5.3	2.7
very	8.0	0.9	7.9	0.9	63	129				16		9.2	0.8
very much	8.7	0.7	8.6	1.0	71	145						8.7	1.5

Source: Project VQS, ROHRMANN 1998

The results indicate:
> for some of the tested VSPLs people differ considerably in their allocation of pertinent intensity levels - see items with high standard deviation sd;
> no significant differences between ratings of context-bound (noise) and context-free presented VSPLs;
> rank order of main VPSLs very similar in CAT, MAG-N and MAG-L scaling results;
> when selecting VSPLs for to-be-labeled 5-point scales, most respondents prefer extreme labels at the end (levels "1" and "5");
> most VSPLs are rated as familiar and easy to understand.

4 Utilization of findings

The results enable the systematic construction of scales measuring the degree of annoyance induced by various noise events/situations and approximating interval scale quality. This can refer to the intensity or frequency or probability of effects; the two other response dimensions (quality and response to statements) are useful as well, e.g., for assessments of noise mitigation. The recommended format is multi-modal, i.e., the scale points should be depicted by a combination of numbers, words perceived as equidistant, and graphical means; thus approximating interval scale level.

Main considerations for choosing a word/expression for a scale point level are:
(1) appropriate position on the dimension to be measured;
(2) low standard deviation;
(3) linguistic compatibility with the other VSPL's of the constructed scale;
(4) sufficient familiarity of the expression;
(5) likelihood of utilization when used in substantive research.
The scale at whole needs to be linguistically coherent and easy to communicate to research participants.

Regarding annoyance scales, a set such as "not/a-little/moderately/quite-a-bit/very" is one possible solutions for a 5-point intensity scale.

The findings of this study are also pertinent for the current discussion within ICBEN-Team-VI (Community Noise) regarding a multi-nationally standardized annoyance scale; cf. Fields 1996, 1998, Guski 1998. The currently considered set, "not-at-all, slightly, moderately, very, extremely" appears to be a reasonable solution (however, there is a rather large gap between level "3" and "4", and using extreme ends can reduce the full usage of a scale).

Finally, it should be noted that the findings from this project are restricted to one language (Australian English) and can not be generalized to English or American or New-Zealand English.

5 The need for further research

To widen the validity scope, three issues seem most relevant: Scaling of labels *within* sets of labels; explicating the impacts of labeling on noise *survey* results; and *cross-cultural* validation (for societal subgroups and across countries). In fact rather complex psycholinguistic research is required to reach the   ultimate aim, namely full international comparability of response scales based on verbal scale point labels..

6 References

Cliff, N. (1959). Adverbs as multipliers. Psychological Review, 66, 27-44.

Fields, J. M. (1996). *Progress toward the use of shared noise reaction questions.* Paper presented at the Internoise 96, Liverpool.

Guski, R. (August 1998).*Personal communication re: "*International Annoyance Scale Project*".*

Hammerton, M. 1986. How much is a large part? Applied Ergonomics, 7, 10-12.

ICBEN -Team VI (1998). *Recommendation for a verbal annoyance scale (draft)* .

Jones, L., & Thurstone, L. L. (1955). The psychophysics of semantics: An experimental investigation. *The Journal of Applied Psychology, 39*, 31-36.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives Psychology, 22*, 1-55.

Moxey, L. M., & Sanford, A. J. (1993). *Communicating quantities: A psychological perspective*. Hillsdale: Erlbaum.

Rohrmann, B. (1978). Empirische Studien zur Entwicklung von Antwortskalen fuer die sozialwissenschaftliche Forschung. *Zeitschrift fuer Sozialpsychologie*, 222-245.

Rohrmann, B. (1985). *Guidelines for social-scientific noise research? Possibilities and problems*. Muenchen: Proceedings of Inter-Noise '85, 105-108.

Rohrmann, B. (in prep. - 1998). *Verbal qualifyers for rating scales: A cross-cultural psychometric study*. Research Report, Univ. of Melbourne, Dept. of Psychology.

Wegener, B. (1983). Category-rating and magnitude estimation scaling techniques: An empirical comparison. *Social Methods and Research, 12*, 31-75.

::