J Dev Phys Disabil (2011) 23:257–265
DOI 10.1007/s10882-010-9224-y

ORIGINAL ARTICLE

Construct Validity, Test-Retest Reliability, and Internal Consistency of the Photo Elicitation Semantic Differential Scale (PESD) in Disability Studies

Bernd A. G. Fellinghauer · Andrea Roth · Katinka Bugari · Jan D. Reinhardt

Published online: 11 January 2011

Abstract The Photo Elicitation Semantic Differential scale (PESD), developed to examine the social perception of disability and attitudes towards people with a disability (PwD), comprises six dimensions: communicativeness, competence, attractiveness, intelligence, industriousness, and popularity. This paper aims to assess the construct validity, test-retest reliability, and internal consistency of the PESD. A longitudinal study with 40 participants of the Swiss general population and 2 (test-retest) * 8 (different photographs) measurements per subject was performed. Construct validity was examined via Principal Component Analysis (PCA), test-retest reliability via the Intraclass Correlation Coefficient (ICC) and a frequency analysis of deviations among test-retest scores, and internal consistency via Cronbach's alpha. PCA extracted two factors corresponding to hard and soft skills for the test and a single factor for the retest. ICCs ranged from 0.44 (industriousness) to 0.60 (intelligence). Deviations between tests exceeding +/-1 were rather rare ranging from 6% (intelligence) to 14% (competence). Cronbach's alpha equalled 0.814 and 0.858 for test and retest, respectively. Summarising, in our study the PESD appears to be a valid

B. A. G. Fellinghauer (✉) · A. Roth · K. Bugari · J. D. Reinhardt
Swiss Paraplegic Research (SPF), Nottwil, Switzerland
e-mail: bernd.fellinghauer@paranet.ch

B. A. G. Fellinghauer
Seminar for Statistics, Swiss Federal Institute of Technology, Zurich, Switzerland

J. D. Reinhardt
Seminar of Health Sciences and Health Policy, University of Lucerne, and SPF, Nottwil, Switzerland

J. D. Reinhardt
ICF Research Branch of WHO FIC CC Germany (DIMDI), SPF, Nottwil, Switzerland

J. D. Reinhardt
IHRS, Ludwig Maximilian University, Munich, Germany

and reliable tool for the examination of the social perception of disability and attitudes towards PwD.

Keywords Disability · Wheelchair · Social perception · Test-retest · Photo elicitation · Semantic differential

Visual impression methods provide an innovative approach to study latent attitudes towards people with a disability (PwD) and stereotypical perception (Goffman 1963). Stereotypical processing of visual clues of impairment may lead to unconscious discrimination against persons with disabilities, which may be counteracted by breaching the latency of the process. In order to assess influence of the visibility of impairment of the appraisal of persons, photo-interviewing or photo-elicitation (Clark-Ibáñez 2004) is a promising approach. Study participants react to photographs showing individuals with and without visible impairments. Generally, photographs activate automatic processing of potentially latent stereotypes more readily than verbal stimuli (Devine 1989).

Semantic differentials (Osgood 1953) are, in turn, well-known scales in attitude research. They use contrary adjectives or nouns between which participants should locate their answer (see Fig. 1). For example, in a study by Mercer et al. (Mercer et al. 1983), a potential female client was photographed in four different conditions: attractive non-disabled, unattractive non-disabled, attractive disabled (in a wheelchair), unattractive disabled. Study participants were asked to rate social attractiveness (semantic differential with positive poles such as 'easy to get along with'), prognosis (semantic differential with positive poles such as 'will be happy', physical attractiveness (semantic differential with positive poles such as 'very attractive'), and personal evaluation (semantic differential with positive poles such as 'reliable').

In a pilot project of the Swiss Paraplegic Research, Reinhardt et al. (Reinhardt et al. 2010) have recently proposed a new scale to examine the social perception of disability and attitudes towards PwD. The pilot study examined whether visual stimuli of impairment activate latent prejudice against disability and whether this connection can be counteracted with priming strategies (here: a cover story told to the participants). Participants were asked to rate 12 photos showing models with mental impairments, physical impairments or no visible impairments. Using mixed models to account for the repeated asking of the same participants and eligible statistical tests, signs of stereotypical processing of visual cues of impairment were found. Based on this newly proposed scale, people with mental disabilities were appraised lower than wheelchair users who in turn were judged worse than people without visible health conditions. This scale, termed Photo Elicitation Semantic Differential scale (PESD), is based on the combination of photo elicitation methods (Clark-Ibáñez 2004) and a semantic differential comprising six dimensions: competence, communicativeness, attractiveness, popularity, industriousness, and intelligence. Along the six dimensions, disabled and non-disabled photo models can be appraised by participants of the general population. The scale is short so that it can be applied with a relatively low burden and in combination with other instruments.

a)

b)

1) Photograph letter: _____ ⇦

	+++	++	+	-	--	---	
competent							incompetent
communicative							uncommunicative
attractive							unattractive
popular							unpopular
industrious							lazy
intelligent							unintelligent
	+++	++	+	-	--	---	

Fig. 1 Part a) of the figure shows the same photo model in a wheelchair and an armchair. Part b) shows the Photo Elicitation Semantic Differential scale (PESD). Ratings range from +++ (best rating) to (—) lowest rating. The scale does not have a middle category

The pilot study examined the scale's potential to test hypotheses. The results indicate the content-wise and predictive validity of the scale. Construct validity and reliability of the instrument have, however, not been explored yet.

The objective of this paper is thus to examine whether the PESD covers one latent dimension (social perception of disability) and whether it is reliable. The specific aims are: 1) to examine the internal validity of the PESD and its eligibility to generate summary-scores, 2) to assess the test-retest reliability, and 3) to check the internal consistency of the scale.

Methods

Design

This study is part of a larger project on visual impressions of impairment and is designed to test psychometric properties of the PESD. The study was planned consistent with a longitudinal design with two (test-retest) * 8 (different photographs) repetitions of measurement per subject.

Sample

The sample was drawn by convenience. 40 participants from the Swiss general population were surveyed in May 2009 and retested in June 2009. Each participant had to appraise the same eight photographs in May and June, yielding for each dimension a total of $40 \times 8 = 320$ pairs of measurement. For communicativeness, one observation is missing resulting in only 319 observed pairs.

As shown in Table 1, 57.5% of the participants were female. Participants were on average 45 years old. Seventy percent of the participants were married or lived in a relationship and over 80% had previous personal contact with PwD.

The Photo Elicitation Semantic Differential Scale (PESD)

The PESD comprises the six dimensions competence, communicativeness, attractiveness, popularity, industriousness, and intelligence, depicted as a semantic differential (Leonard and Crawford 1989; Osgood 1953) with six ranks from --- (worst) to +++ (best). The PESD does not have a middle category. The dimensions were chosen based on existing studies (Bonfranchi et al. 2002) and should represent properties relevant for social success in modern society. Figure 1 provides: a) two example photographs, and b) a depiction of the semantic differential.

Table 1 Socio-demographics

Variable	Value	Mean (sd)/percent
Age (years)		45.01 (18.8)
Sex	Female	57.50%
Marital status	Single	25%
	Married or in a relationship	70% (=45%+25%)
	Divorced or broken up	5%
Highest level of education	High (university, applied university)	45%
	Middle (job training, A-levels)	50%
	Low (secondary school)	2.50%
Previous contact to PwD	yes	81.30%

The table shows the mean/percent values of the most important socio-demographic variables. Values in brackets denote standard deviations. The data was available for all 40 participants.

Procedure

Eight photo models (four paraplegics and four persons with no visible disability) were photographed twice: once in an armchair and once in a wheelchair. The photos were taken by a professional photographer. Informed consent was secured from the models and proxies who gave their written permission to use the photographs in scientific studies and publications.

Photographs were split into two sets such that each photo model was shown to the study participants only once, either in a wheelchair or in an armchair. Each set was organized such that exactly half of the photo models in a set have a spinal cord injury. Furthermore, the two sets were similar in age and identical in sex distribution. Each of the 40 study participants was randomly allocated to one of the two sets of photographs resulting in two equally-sized groups. Each group was then asked to appraise their set of eight photographs in the aforementioned six dimensions. Photographs were shown in randomized order according to predefined randomization lists at both points of measurement.

Data Analysis

Construct validity was examined with a Principal Component Analysis (PCA) with orthogonal rotation (varimax) calculated for both measurement points. Factors with eigenvalues greater than one were extracted (Kaiser criterion (Kaiser 1960)).

Test-retest reliability between the two measurements was assessed using the Intraclass Correlation Coefficient ICC (3,1) (Shrout and Fleiss 1979), ranging from 0 to 1. Higher values indicate better test-retest reliability. A frequency analysis provided an overview of the size of the deviations between the two measurement points.

Internal consistency of the instrument was assessed using Cronbach's alpha (Cronbach 1951), ranging from 0 to 1. A higher value indicates a more consistent instrument.

All analyses were performed using SPSS 14.

Results

PCA revealed a two factor solution (see Table 2) for the first measurement point and a one factor solution for the second measurement point (not shown). For the first measurement point, variables representing soft skills (communicativeness, popularity, and attractiveness) loaded on one factor and hard skills (intelligence, competence and industriousness) on the other (see Table 3). Each factor explains roughly 35% of the variance of the variables in the rotated solution (around 70% in total (see Table 2); explained variance of rotated solution not shown). The second measurement's single extracted factor explains about 53% of the items' variance (not shown) and is comprised of items from both the soft skills and hard skills domain (not shown).

For test-retest-reliability, ICCs range from 0.44 (industriousness) to 0.60 (intelligence). The remaining dimensions all exceed 0.5 (see Table 4). From the

Table 2 Principal components analysis

Factor	Unrotated PCA solution			Unrotated factors extracted		
	Eigenvalues	Percent of variance explained	Cumulated percent	Eigenvalues	Percent of variance explained	Cumulated percent
1	3.185	53.08	53.08	3.185	53.08	53.08
2	1.017	16.95	70.03	1.017	16.95	70.03
3	0.572	9.53	79.56			
4	0.522	8.70	88.26			
5	0.377	6.28	94.54			
6	0.328	5.46	100.00			

The table shows the results from a principal components analysis at the first point of measurement. The original (unrotated) solution and the corresponding eigenvalues and explained variances are given. From a varimax rotation two factors were extracted (Kaiser Criterion) which, as can be seen in Table 3, correspond to hard skills and soft skills.

frequency distribution of the difference "appraisal at measurement point two minus appraisal at measurement point one" (see Table 5), a spectrum of −3 to +4 units of deviation was found. With almost 50% identical measurements, popularity is the most stable dimension, closely followed by intelligence (48%) and competence (47%). The least stable dimension is communicativeness (37%) followed by attractiveness (39%) and industriousness (42%). Deviations by either +1 or −1 unit are in the magnitude of 43% (popularity) to 51% (attractiveness). Deviations exceeding +/−1 are rather rare ranging from 6% (intelligence) to 14% (competence).

Internal consistency for measurement points one and two, expressed by Cronbach's alpha, equals 0.814 and 0.858, respectively. The internal consistency of the potential two 3-item subscales is equal to a Cronbach's alpha of 0.79 (hard skills) and 0.72 (soft skills), respectively, for the first measurement point, and 0.80 (hard skills) and 0.80 (soft skills) for the second measurement point.

Table 3 Extracted factors

	Component	
	Hard skills	Soft skills
Competence	0.704	0.457
Communicativeness	−0.021	0.866
Attractiveness	0.425	0.648
Popularity	0.342	0.747
Industriousness	0.863	−0.01
Intelligence	0.788	0.325

The table shows the individual factor loadings of the six dimensions on the two varimax-rotated factors "hard skills" and "soft skills" extracted from a principal components analysis at the first point of measurement (Kaiser Criterion).

Table 4 Intraclass correlation coefficient

Dimension	ICC(3,1)	95% lower bound	95% upper bound	Valid pairs
Competence	0.56*	0.48	0.63	320
Commucativeness	0.51*	0.43	0.59	319
Attractiveness	0.58*	0.49	0.66	320
Popularity	0.53*	0.45	0.61	320
Industriousness	0.44*	0.35	0.53	320
Intelligence	0.60*	0.53	0.67	320

The table shows the Intraclass Correlation Coefficient (ICC(3,1)) and the estimated boundaries for the six dimensions. Valid pairs refer to the 320 measurement pairs (40 persons rate eight photographs at two time points). A * indicates a significant deviation from zero ($p<0.01$).

Discussion

The PESD shows a differentiation into two latent dimension representing hard and soft skills for the first measurement point. Habituation effects seem, however, to lead to less differentiated appraisals in the second measurement point resulting in a single factor comprising of items from both the hard and soft skills domain. The PESD showed moderate to good test-retest properties, as most ICCs exceed 0.5. This is an adequate value for an empirical study. Also, frequency analysis further supports our findings from ICC analysis in that it reveals only slight deviations between the two measurements which rarely exceed one scale unit. It is interesting to note that some dimensions, e.g.

Table 5 Frequencies

Variable	Frequency	−3	−2	−1	0	1	2	3	4	Total
Competence	Absolute	1	8	49	151	95	13	3	0	320
	Relative	0.31	2.5	15.31	47.19	29.69	4.06	0.94	0	100
Communicativeness	Absolute	3	18	80	119	76	18	5	0	319*
	Relative	0.94	5.63	25	37.19	23.75	5.63	1.56	0	99.69*
Attractiveness	Absolute	0	7	56	126	107	18	6	0	320
	Relative	0	2.19	17.5	39.38	33.44	5.63	1.88	0	100
Popularity	Absolute	0	7	64	159	75	11	4	0	320
	Relative	0	2.19	20	49.69	23.44	3.44	1.25	0	100
Industriousness	Absolute	1	16	67	135	79	19	2	1	320
	Relative	0.31	5	20.94	42.19	24.69	5.94	0.63	0.31	100
Intelligence	Absolute	1	7	75	155	72	9	1	0	320
	Relative	0.31	2.19	23.44	48.44	22.5	2.81	0.31	0	100
Total	Absolute	6	63	391	845	504	88	21	1	1919
	Relative	0.3	3.4	20.4	44	26.4	4.5	1	0	100

The table shows the frequency of deviations among measurements for the two time points. A positive (negative) value refers to an increase (decrease) of the average score in the second measurement. Zero indicates no such deviation. The * indicates a missing measurement.

intelligence show noticeably better test-retest reliability than other items (e.g. industriousness). Possibly, some of the investigated dimensions are more intuitive to the participants and so their ratings on a retest will be closer to the first testing.

Internal consistency of the PESD is promising as both measurements exceed 0.8. The increase in Cronbach's alpha, found for the second measurement point, might be attributed to habituation effects. The lower alpha values of the 3-item subscales (relative to the total scale) can be seen as an artifact to the smaller item pool of the subscales, as Cronbach's alpha uses the Spearman-Brown formula to upshift the reliability corresponding to the number of items in the scale (Brown 1910).

In our study, the scale seems to be suitable for the measurement of (latent) attitudes towards persons with disabilities along these dimensions. It is important to note, however, that we only used photographs of persons with physical disabilities in this study. Corresponding psychometric properties for the PESD in studies with photo models with mental disabilities have to be established in a future study. Also, the study population was drawn by convenience from the Swiss general population and might not be representative for international comparisons.

In future studies with a focus on the photographic appearance of the person, for example in the context of a job interview or social networks, the PESD might allow for a more indirect measurement of attitudes towards disabled persons than a classical survey approach (e.g., the Interaction with Disabled Persons scale (Gething and Wheeler 1992)). Future research will be necessary to judge whether the identified underlying latent constructs "hard skills" and "soft skills" can be reproduced for different study populations. The PESD could then ultimately be used to develop summary scores of attitudes towards PwD.

Conclusion

The PESD appears to measure two latent dimensions, but habituation seems to lead to a more one-dimensional perception. This study yielded good test-retest reliability and internal consistency of the PESD. The scale might thus be a useful tool to study social perception of disability and attitudes towards PwD.

Acknowledgements The authors like to thank all participants of this study. In particular, we would like to thank all photo models. Furthermore, we thank all reviewers and the editorial office for their valuable feedback.

Declaration of interest The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Bonfranchi, R., Mayer, E., & Rupp, D. (2002). *Visueller Eindruck—geistige Behinderung—gesellschaftliche Bedeutung*. Oberhausen: ATHENA-Verlag.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3*, 296–322.

Clark-Ibáñez, M. (2004). Framing the social world with photo-elicitation interviews. *The American Behavioral Scientist, 47*(12), 1507–1527.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.
Devine, P. G. (1989). Stereotypes and prejudice: their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 6–18.
Gething, L., & Wheeler, B. (1992). The interaction with disabled persons scale: a new Australian instrument to measure attitudes towards people with disabilities. *Australian Journal of Psychology, 44* (2), 75–82.
Goffman, E. (1963). *Stigma. Notes on the management of spoiled identity (First Touchstone Edition ed.)*. New York: Simon & Schuster.
Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141–151.
Leonard, R., & Crawford, J. (1989). Two approaches of seeing people with disabilities. *Australian Journal of Social Issues, 24*(2), 112–125.
Mercer, J., Andrews, H., & Mercer, A. (1983). The effects of physical attractiveness and disability on client ratings by helping professionals. *Journal of Applied Rehabilitation Counseling, 14*, 41–45.
Osgood, C. E. (1953). *Method and theory in experimental psychology.* Oxford University Press.
Reinhardt, J. D., Ballert, C. S., Fellinghauer, B., Lötscher, A., Gradinger, F., Hilfiker, R., et al. (2010). Visual impressions of disability and the social construction of the person. A randomized controlled field experiment using photo elicitation. *Disability and Rehabilitation, Early online*, 1-12.
Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 2*, 420–428.