

Paper and pencil versus web

Does the format affect the psychometric properties
of a questionnaire measuring adolescent health?

Curt Hagquist

**Centre for Research on Child and Adolescent Mental Health
Karlstad University, Sweden**

Michael Hansen

**Department of Education and Special Education
University of Gothenburg, Sweden**

**5th Conference of the International Society for Child Indicators
From Welfare to Well-being: Child indicators in research, policy & practice.
2 – 4 September 2015, University of Cape Town, South Africa**



Outline

- Introduction and background
- Differential Item Functioning (DIF)
- Comparisons of paper and pencil vs web questionnaires
 - item level
 - person measurement
- Conclusions



Introduction

Web-based questionnaires are increasingly being used to collect survey data also among adolescents. If, and to what extent, the choice of format for a questionnaire affects the responses from the participants is still not definitively answered.



Sample of previous studies: 1

“The Internet group had more adolescents reporting that they have a sufficient number of friends than those in the paper mode (p .01). The Internet mode received more favorable evaluations than P&P. The multiple items per screen format was favored over the one item per screen format on perceived speed of the administration mode.”

Mangunkusumo RT, Moorman PW, Van Den Berg-de Ruitter AE, Van Der Lei J, De Koning HJ, Raat H.(2006) Internet-administered adolescent health questionnaires compared with a paper version in a randomized study. *J Adolesc Health*, 36(1), 70.e1-6.



Sample of previous studies: 2

“...found that mode of administration had no significant effect on adolescents’ responses, whereas for several questions about feelings/affective states, more socially desirable responses were found in the paper and pencil format than in the computer format.”

Vereecken CA, Maes L. (2006). Comparison of a computer-administered and paper-and-pencil-administered questionnaire on health and lifestyle behaviors. *J Adolesc Health*, 38(4), 426-32.



Sample of previous studies: 3

“The findings suggest that in a controlled school setting, web-based administration of health indicators yields almost the same results as paper-and-pencil administration.”

van de Looij-Jansen & de Wilde. (2008). Comparison of web-based versus paper-and-pencil self-administered questionnaire: Effects on health indicators in dutch adolescents. *Health Service Research*, 43(5), 1708-1721.



Sample of previous studies: 4

“Findings included less missing data for the paper-and-pencil condition (1.5% vs. 5.3%, 4.4%, 6.4%; $p < .001$), less perceived privacy and anonymity among respondents for the in-class Web conditions, and a lower response rate for the “on your own” Web condition than for in-class administration by either mode (28.0% vs. 91.2%, 90.1%, 91.4%; $p < .001$). Although Web administration might be useful for some surveys, these findings do not favor the use of a Web survey for the YRBSS.”

Denniston M, Brener N, Ross J, et al. (2010). Comparison of paper-and-pencil versus Web administration of the Youth Risk Behavior Survey (YRBS): Participation, data quality, and perceived privacy and anonymity. *Computers In Human Behavior* [serial online]. January 1, 2010;26(Advancing Educational Research on Computer-supported Collaborative Learning (CSCL) through the use of gStudy CSCL Tools):1054-1060.



Sample of previous studies: 5

“No statistically significant differences were found for self-reported risk across modes of administration. Students completing the Web-based version of the survey were four times more likely to skip an item.”

David L. Wyrick and Lloyd Bond (2011) Reducing Sensitive Survey Response Bias in Research on Adolescents: A Comparison of Web-Based and Paper-and-Pencil Administration. *American Journal of Health Promotion*, 25(5), 349-352.



Sample of previous studies – 6

“We conduct differential item functioning (DIF) analysis to assess whether item-response probability is different based on survey format. Results demonstrate that young people completing the M&MS on paper have lower scale-level overall scores. However, DIF analyses indicate that this difference is not explained by item-level probabilities.”

Patalay, P., Deighton, J., Fonagy, P., & Wolpert, M. (2015). Equivalence of paper and computer formats of a child self-report mental health measure. *European Journal of Psychological Assessment*, 31(1), 54-61.



Purpose

The purpose of the present study is to psychometrically examine the concordance between responses from a web-based questionnaire and a paper and pencil questionnaire. For the purpose of this study eight items from The PsychoSomatic Problems (PSP)-scale were subjected to a psychometric analysis based on Rasch Measurement Theory (RMT).



Data Collection

Data were collected during the first half of 2014 by Statistics Sweden among students in grade 9 (15-16 years old). A questionnaire about school and living conditions was sent by regular post to a nationwide sample of individuals participating in a longitudinal cohort study run by the University of Gothenburg.

The participants could choose whether to complete the questionnaire with a paper and pencil version or a web version. 3904 students (86 %) preferred the former version and 615 students (14 %) the latter one.



Use of questionnaire format distributed by choice of upper secondary school

Choice of upper secondary school				
	Public school in the neighborhood	Other public school	Independent school	No choice
n =	1273	2024	1064	93
Paper and pencil	86,7%	86,5%	84,5%	90,3%
Web	13,3%	13,5%	15,5%	9,7%



Instrument

The PsychoSomatic Problems scale (PSP)

- “had difficulty in concentrating”
- “had difficulty in sleeping”
- “suffered from headaches”
- “suffered from stomach aches”
- “felt tense”
- “had little appetite”
- “felt sad”
- “felt giddy”

The response categories for all of these items, which are in the form of questions, are “always” “often” “sometimes” “seldom” and “never”.

The recall period concerned the last six months.



Analysis

The main focus of the analysis was on Differential Item Functioning (DIF), i.e. if the PSP-items worked in the same way for students completing a web questionnaire and a paper and pencil questionnaire respectively. The Rasch analysis was performed with the software RUMM2030.



Differential Item Functioning



DIF – a special case of lack of invariance

“A comparison of a pair of items across sexes. More elaborate tests of fit that incorporate the invariance across both the total score groups and across different classes of persons can be constructed.”

Andrich (1988). Rasch Models for Measurement, p 72.



No DIF definition

For the same values of the latent variable, the expected value of a response to an item is identical for members from any group of individuals.



DIF across a priori specified sample groups

- Uniform DIF
- Non uniform DIF



Uniform group DIF

For the same values of the latent trait, the expected value of a response to an item from members from one group of individuals is different from members of another group – and the differences are the same along the whole latent trait [= parallel slopes].



Non uniform group DIF

For the same values of the latent trait, the expected value of a response to an item from members from one group of individuals is different from members of another group – and the differences vary along the latent trait [= non parallel slopes].



The method used for analysis of DIF

Two-way analysis of variance of standardised residuals.



Resolving DIF

- Both uniform DIF and non uniform DIF may be accounted for by split of the DIF-items. This implies that each of the sample group specific items may statistically fit the Rasch model.
- However, invariance is not retained, which manifested by different item parameters across sample groups for the same item.
- Resolving DIF may be justified if the source of DIF can be assumed to be a result of different proneness to endorse a certain item, or some other aspect irrelevant to the content of the variable and therefore deemed dispensable.



Results: Targeting



Person-Item Threshold Distribution

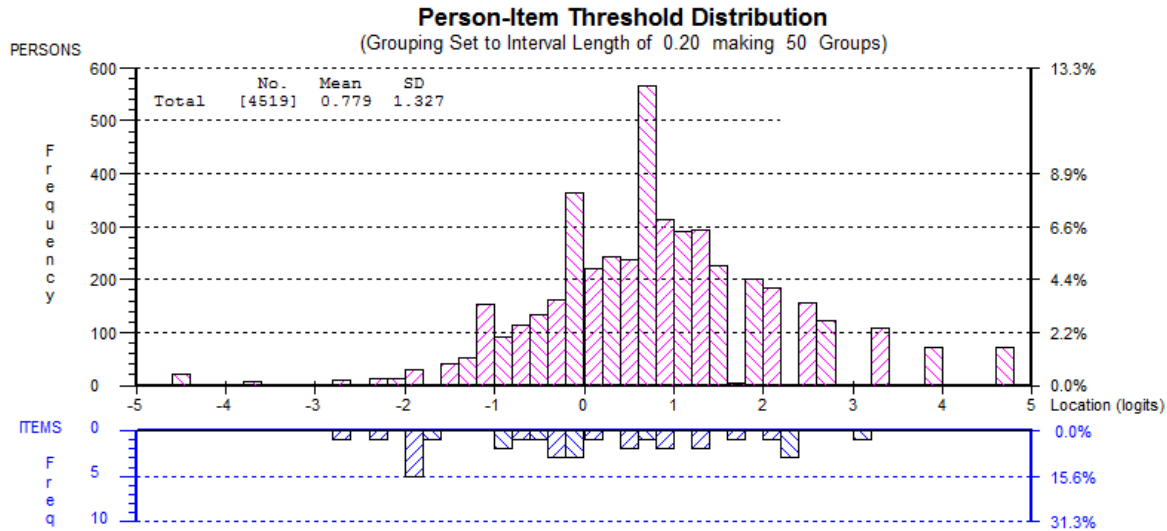
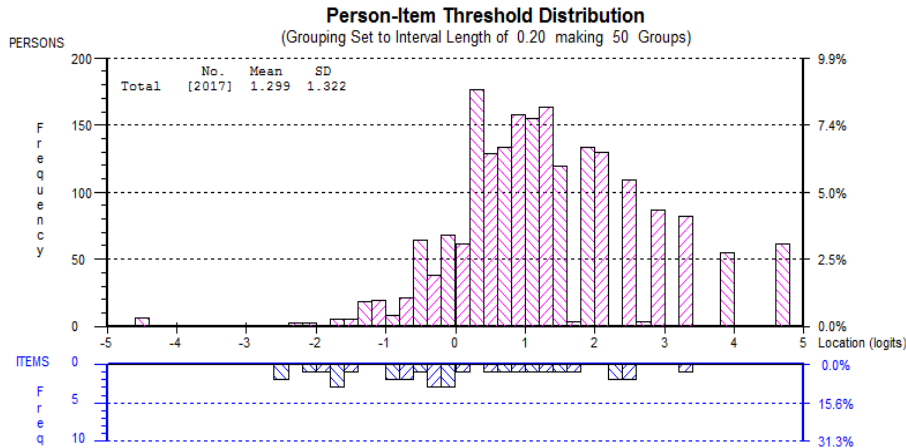


Figure 1. Person-Item Threshold Distribution.

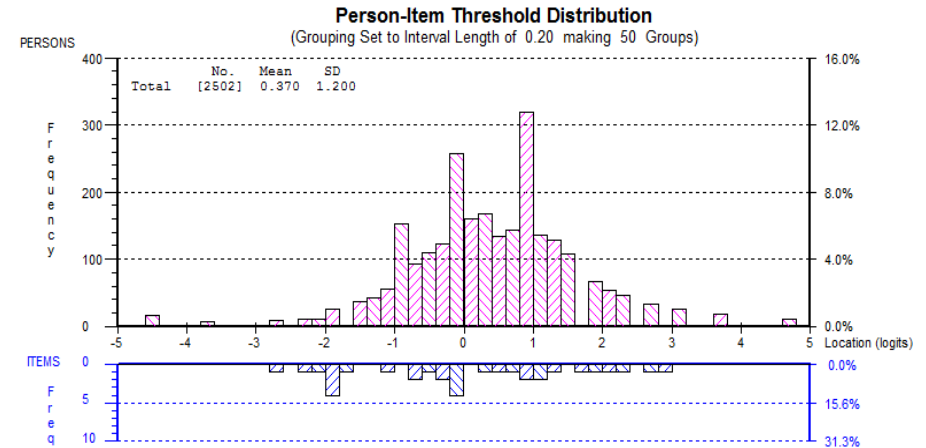


Person-Item Threshold Distribution

2a



2b



Figures 2 a-b. Person-Item Threshold Distribution for boys and girls.



Results: Differential Item Functioning



Differential Item Functioning by SexFormat

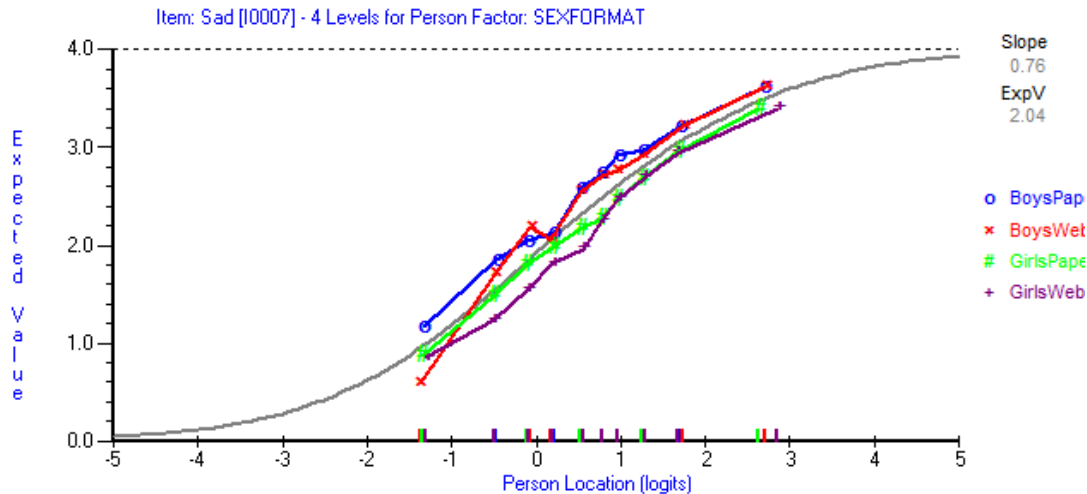
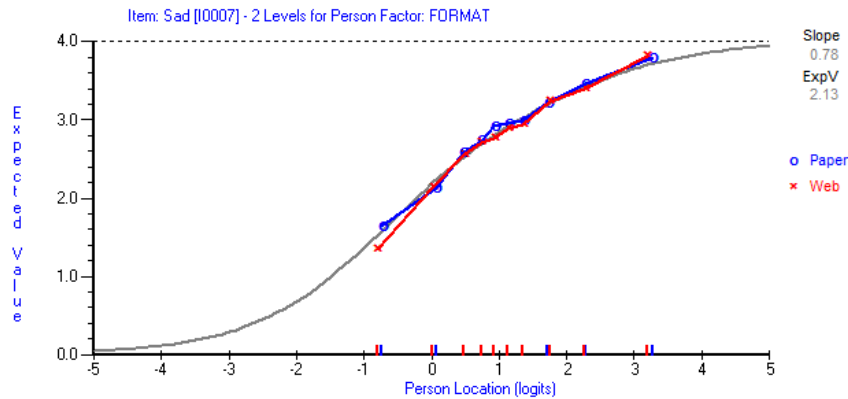


Figure 3. Differential Item Functioning by SexFormat for item Sad.

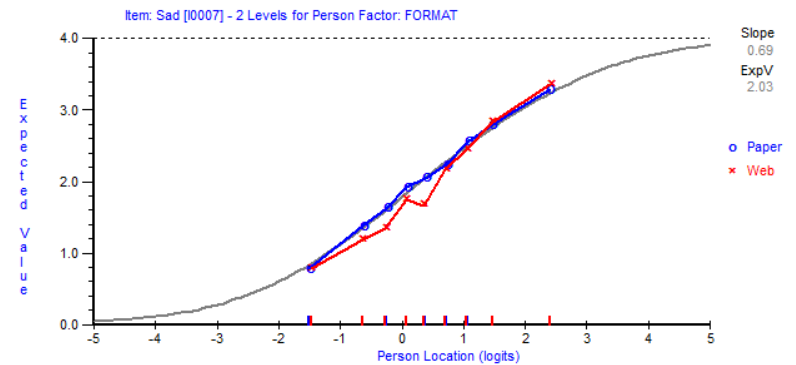


Differential Item Functioning by Format

4a



4b



Figures 4 a-b. Differential Item Functioning by Format for item Sad for boys and girls.



Results: Person Measures



Means of person logit values

	SexFormat			Format – Boys			Format – Girls		
	Number	Mean	sd	Number	Mean	sd	Number	Mean	sd
BoysPaper	1 680	1.331	1.25				[2 224]	0.407	1.18
BoysWeb	337	1.176	1.43				[278]	0.071	1.32
GirlsPaper	2 224	0.392	1.19	[1 680]	1.325	1.29			
GirlsWeb	278	0.056	1.32	[337]	1.167	1.47			



Why differ the person mean values between the two questionnaire formats?

- Selection, e.g. differences in the composition of the two groups, socio-demographic or others?
- Differences in the proneness to disclose information on sensitive questions?
- ...or?



Summary of the results

There is no or only minor DIF with respect to choice of questionnaire format. Given the same location on the latent variable the expected value for each item is the same for students completing a web-questionnaire as for students completing a paper and pencil questionnaire. The reliability measured by a person separation is good (~ 0.85) for the entire sample as well as for each of the format sub groups.

There are no reversed latent thresholds, indicating that the categorization of the items works as intended. In the analyses of the entire data set and in both subsets the item “felt sad” shows gender DIF implying that given the same location on the latent trait girls score higher frequency of complaints than boys.



Conclusions

The psychometric properties of the PSP-scale don't seem to be affected by the choice of format for completing the questionnaire. Regardless of format issue attention needs to be paid to the item "felt sad". Resolving this DIF-item by splitting it into two gender-specific items is an option to consider in further analyses.

Also other aspects of the choice of questionnaire format need to be examined, e.g. the impact on the response rates.



THANKS!
curt.hagquist@kau.se
www.kau.se/en/cfbuph

