

Presentation of a Single Item Versus a Grid: Effects on the Vitality and Mental Health Subscales of the SF-36v2 Health Survey

Mario Callegaro, Jeffrey Shand-Lubbers, J. Michael Dennis: Knowledge Networks

Introduction

There is a lot of debate whether questions should be presented on a grid or on a single item per screen.

From just an operational point of view, grids take less time to complete, which should decrease response burden, but new research shows that respondents seem to prefer a single item per screen (Garland, P., 2009; Thorndike, F. P. et al, 2009).

From a measurement point of view, grids pose numerous issues:

- **Higher item nonresponse** (Couper, Traugott, Lamias, 2001; Toepel, Das and van Soest, forthcoming)
- **Higher Cronbach's alpha** (Tourangeau, Couper & Conrad, 2004; Toepel, Das and van Soest, forthcoming)
- **Higher item non-differentiation** (Tourangeau, Couper & Conrad, 2004)
- **Sometimes higher measurement error** (Peytchev, 2006) **but not always** (Thorndike, F. P. et al, 2009)

Previous research on the SF-36 found contrasting results: Bell, Mangione & Kahn (2001) did not find difference in alpha when questions were placed on a grid (compared to single item per screen) in a web survey while Iglesias, Birks and Torgerson (2001) found higher alpha level when questions were placed in a vertical sequence in comparison to a grid in a paper instrument.

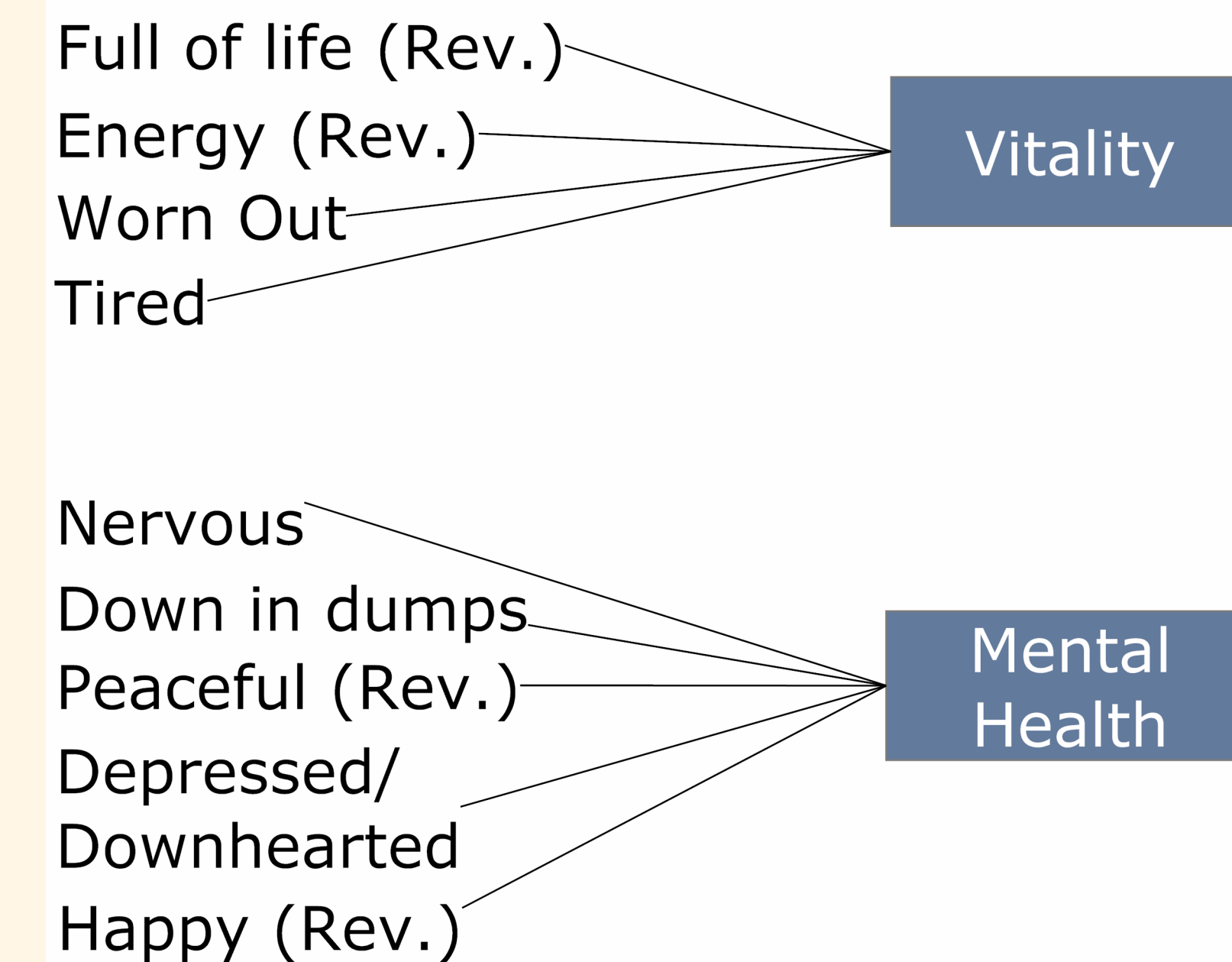
Although every researcher is faced with deciding if organizing the questions in a single or multiple-item per screen, the amount of research on this topic is "surprisingly light" (Couper, 2008, p. 191).

Items used in the experiment

In this experiment we are testing the Vitality and Mental Health subscales of the SF-36v2®.

The SF-36v2 Health Survey asks 36 questions to measure functional health and well-being from the patient's point of view. It is called a generic health survey because it can be used across age (18 and older), disease, and treatment group, as opposed to a disease-specific health survey which focuses on a particular condition or disease.

Measurement model



Note: this is not the full SF-36v2 model. SF-36v2® is a trademark of the Medical Outcomes Trust

Study design

A sample of 2,500 KnowledgePanel® respondents were randomly assigned to one of 5 experimental conditions:

- Group 1: Standard grid
- Group 2: Shaded grid
- Group 3: One item per screen with horizontal response options
- Group 4: One item per screen with vertical response options
- Group 5: One item per screen with vertical shaded response options

Completion rate was of 73.4%. Respondents who completed the study either on a MSNTV or on a iPhone were excluded because the survey was not optimized for that type of browser. Approximately 290 cases per group were used in the analysis.

Hypotheses

- Grid conditions should elicit higher item nonresponse than single item presentation
- Single item conditions should obtain higher satisfaction than grid presentations
- Grid answers should be "less consistent" than single item answers because of the reversed direction of 4 items

Analysis

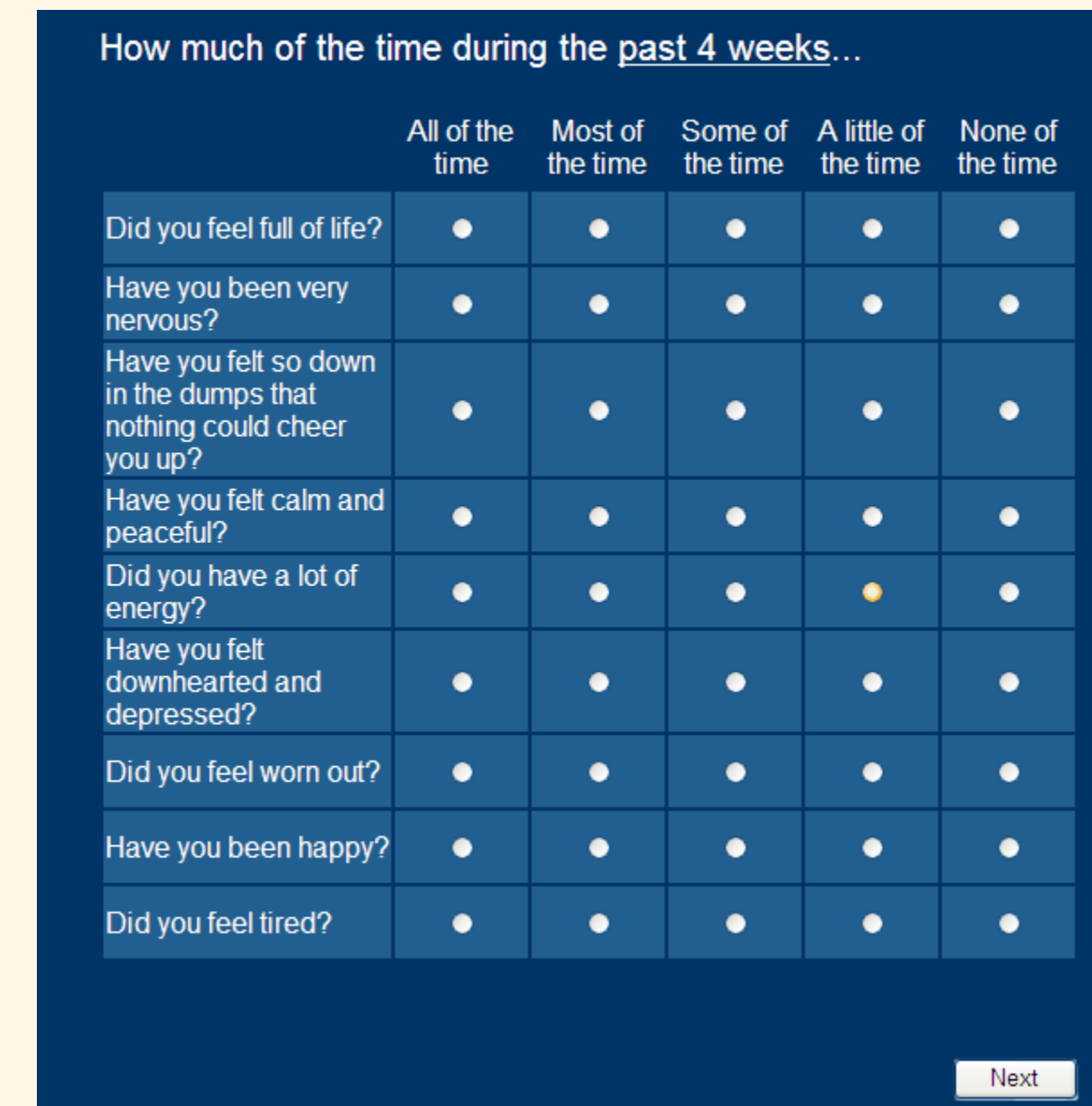
Compare the two subscale scores across conditions

Compare Cronbach's alpha across conditions

Compare difficulty and enjoyment ratings across conditions

Screenshots

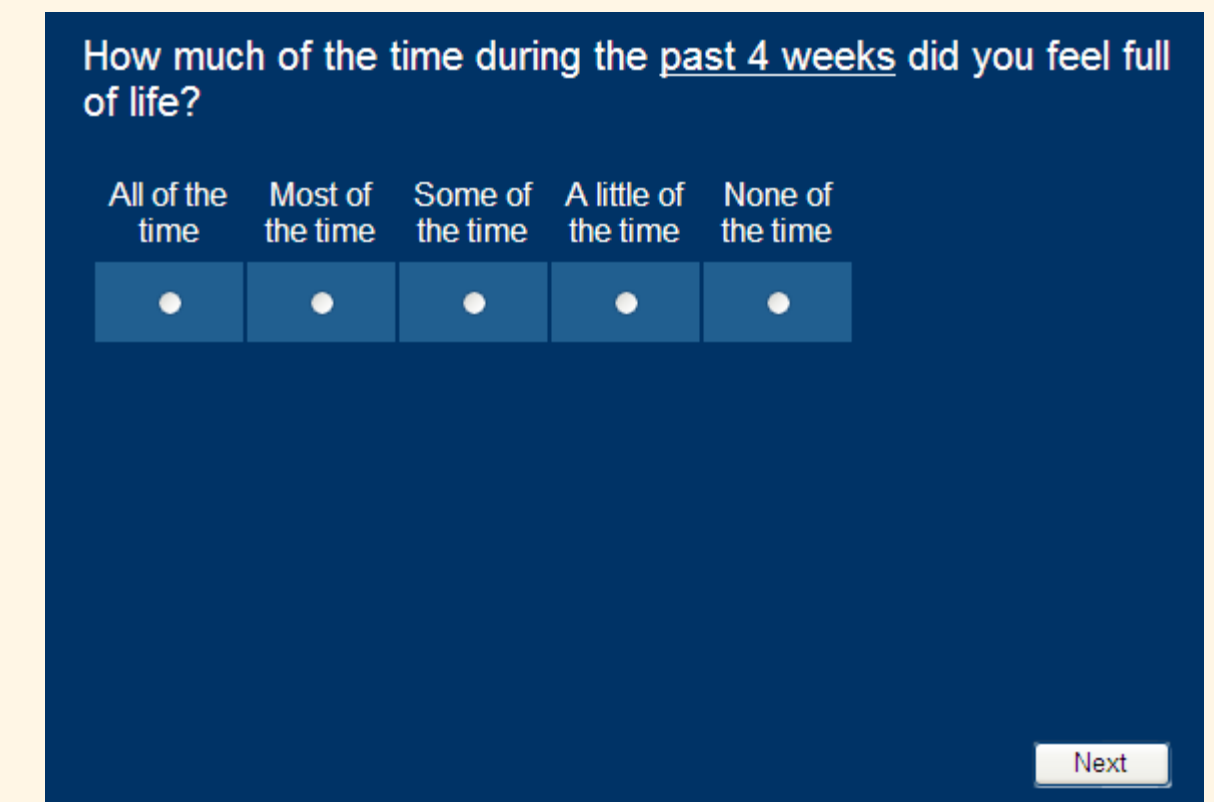
Group 1: Standard grid



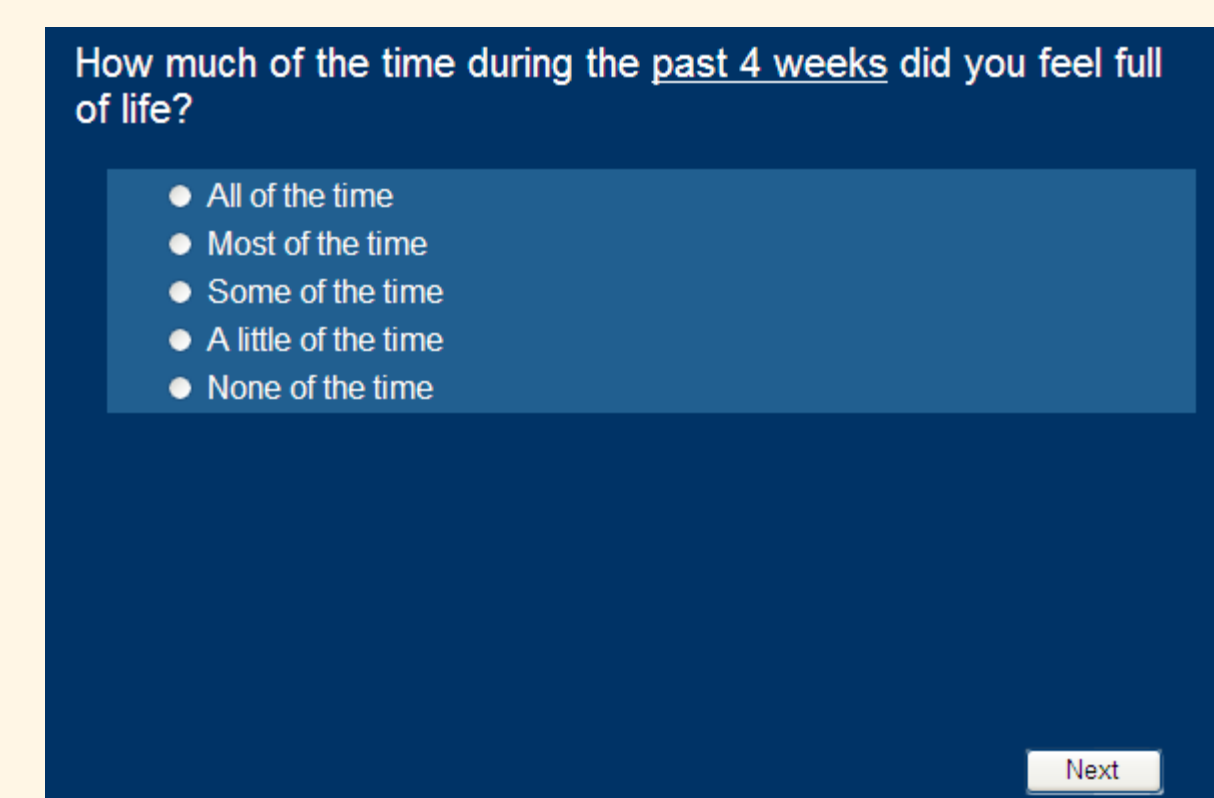
Group 2: Shaded grid



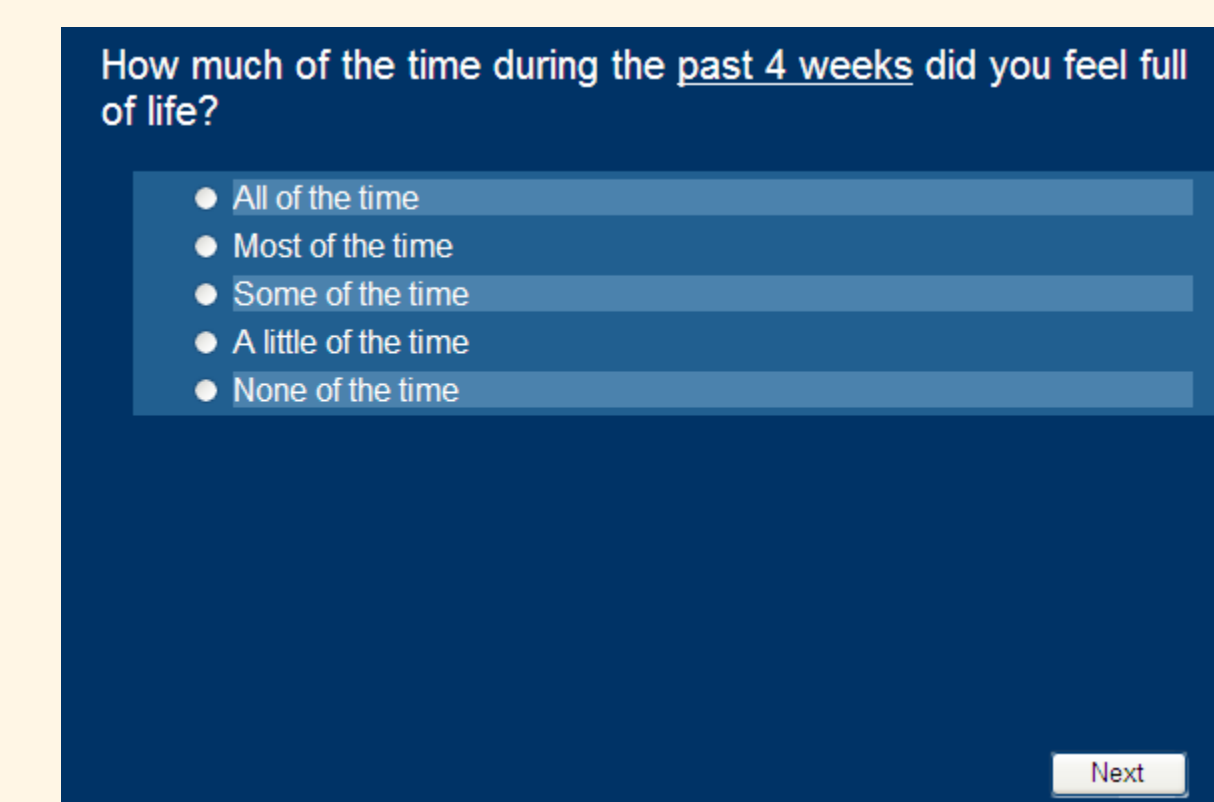
Group 3: One item horizontal response options



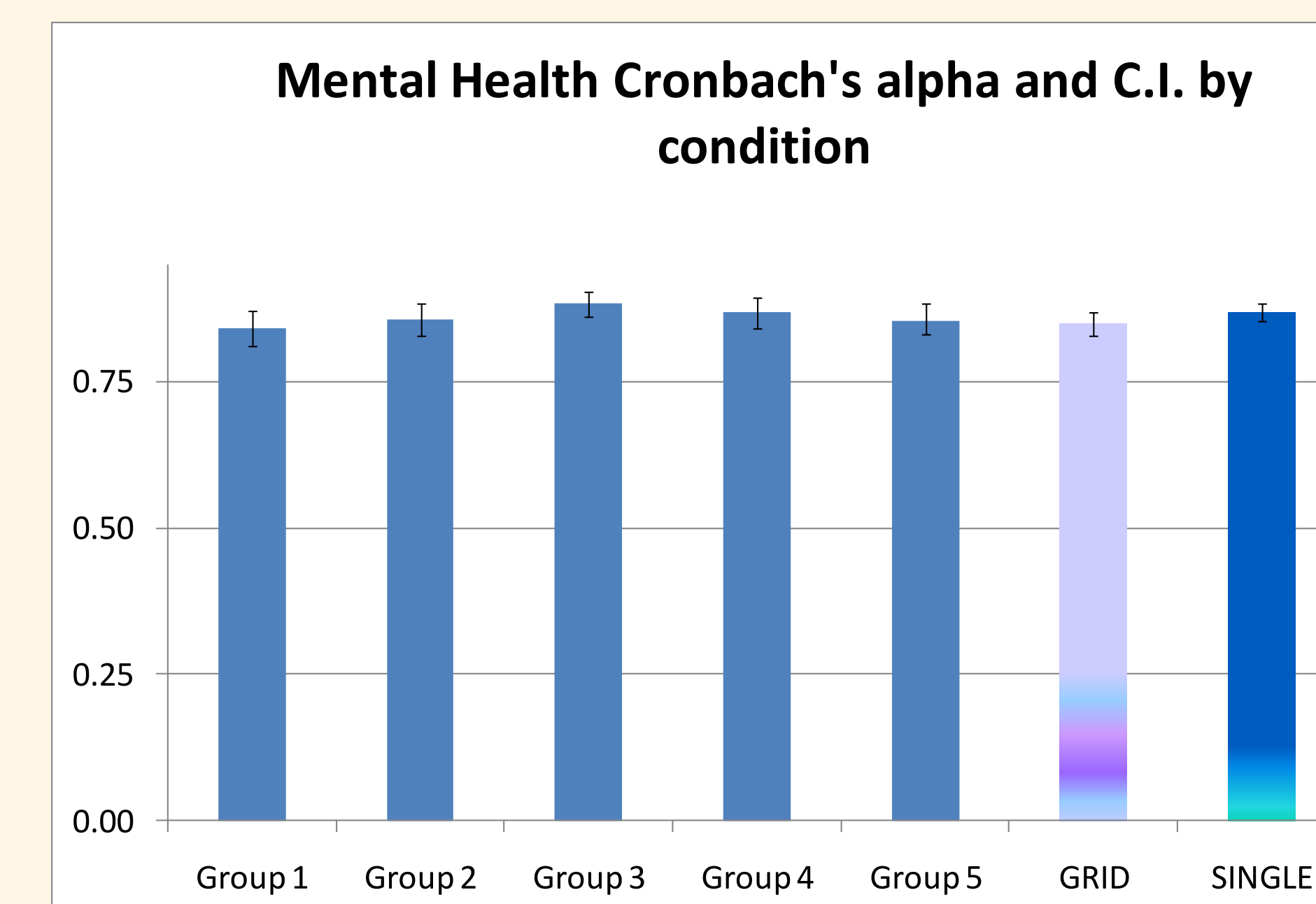
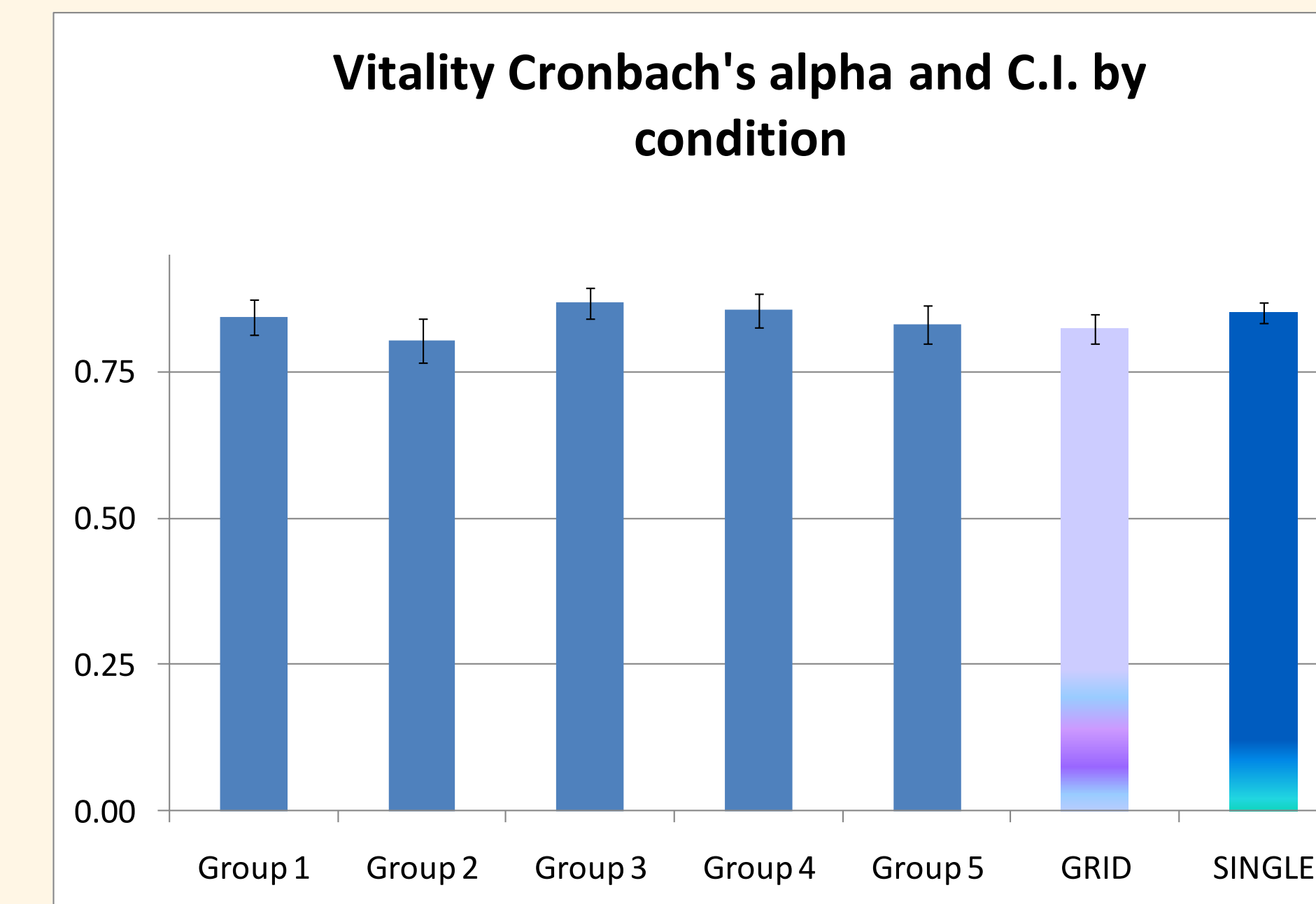
Group 4: One item vertical response options



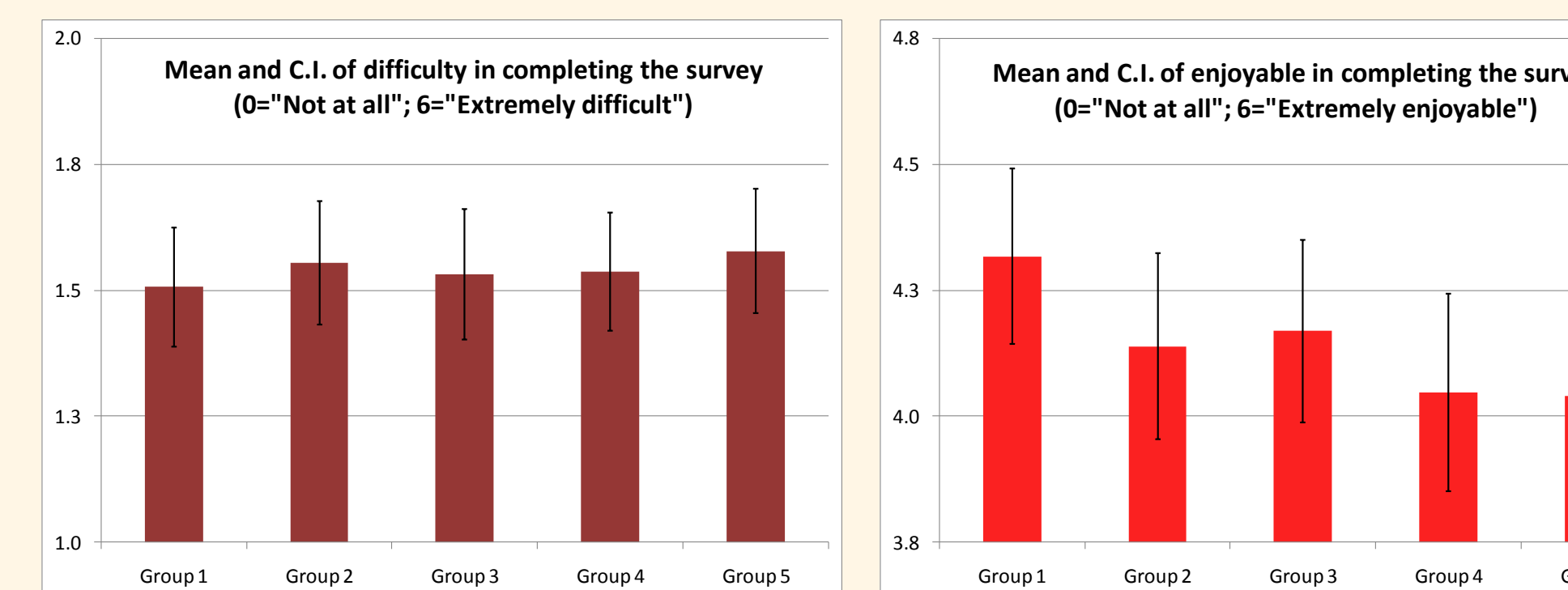
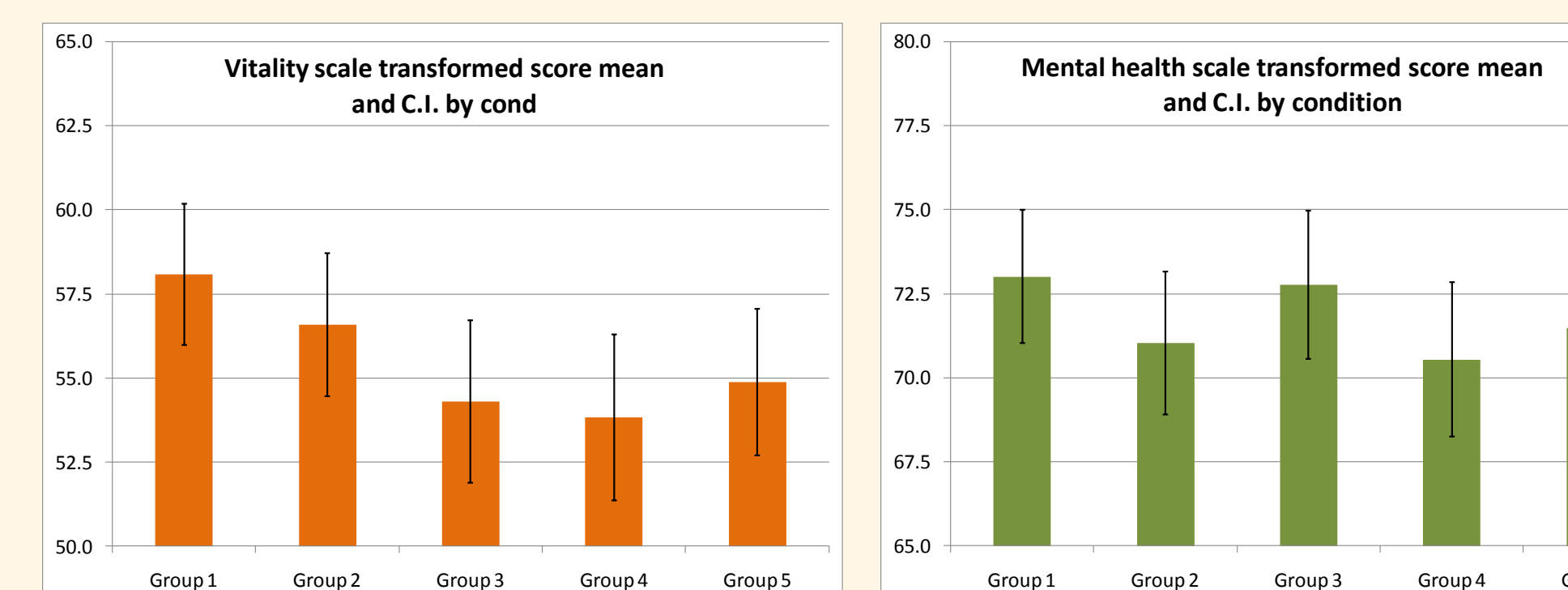
Group 5: One item vertical shaded resp. opt.



Results



C.I. computed using Duhachek & Iacobucci (2004) formula.



Item nonresponse

	Grids	Single item
Mean	0.0077	0.0048
SD	0.0640	0.0439

t=1.011; P= 0.312

Correlation between the sum of the scores for the reversed items and the sum of the scores for the non-reversed items by groups and by scale.

Pearson's r	Grids	Single item
Vitality	.550	.594 ^{ns}
Mental health	.629	.671 ^{ns}

Conclusions

Items on a grid appear to increase measurement error although the differences did not reach statistical significance.

Presenting items one by one allows respondents to focus more on the question and the response options than when presented together as a grid.

It seems that on a grid it is much easier to "miss" the meaning of the items, resulting in more inconsistencies than when evaluating one item per screen (see results of Person's r).

We plan to do more analysis using Confirmatory Factor Analysis and Structural Equation Modeling.

Selected References

Bell, D. S., Mangione, C. M., & Khan, C. E. (2001). Randomized testing of alternative survey formats using anonymous volunteers on the world wide web. *Journal of the American Medical Informatics Association*, 8, 616-620.

Couper, M. P. (2008). *Designing effective web surveys*. Cambridge: Cambridge University Press.

Couper, M. P., Traugott, M. W., & Lamias, M. J. (2001). Web survey design and administration. *Public Opinion Quarterly*, 65, 230-253.

Duhachek, A. & Iacobucci, Dawn (2004). Alpha's standard error (ASE): An accurate and precise confidence interval estimate. *Journal of Applied Psychology*, 89, 792-808.

Garland, P. (2009). *Grid test summary*. Survey Sampling International.

Iglesias, C. P., Birks, Y. F., & Torgerson, D. J. (2001). Improving the measurement of quality of life in older people: the York SF-12. *QJM*, 94, 695-698.

Peytchev, A. (2006). Web survey design: Effect of layout on measurement error. In C. van Dijkum, J. Blasius & C. Durand (Eds.), *Recent developments and applications in social research methodology: Proceedings of the RC33 Sixth International Conference on Social Science Methodology, Amsterdam 2004 [CD-ROM]*. Opladen & Farmington Hills: Barbara Budrich.

Thorndike, F. P., Calbring, P., Smyth, F. L., Magee, J. C., Gonder-Frederick, L., Ost, L.-G., et al. (2009). Web-based measurement: Effect of completing single or multiple items per webpage. *Computers in Human Behavior*, 25, 393-401.

Toepel, V., Das, M., & van Soest, A. (forthcoming). Design of web questionnaires: A test for number of items per screen. *Field Methods*.

Tourangeau, R., Couper, M. P., & Conrad, F. G. (2004). Spacing, position, and order interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly*, 68, 368-393.