

# Investigating measurement equivalence of visual analogue scales and Likert-type scales in Internet-based personality questionnaires

Tim Kuhlmann<sup>1</sup> · Michael Dantlgraber<sup>1</sup> · Ulf-Dietrich Reips<sup>1</sup>

Published online: 27 January 2017  
© Psychonomic Society, Inc. 2017

**Abstract** Visual analogue scales (VASs) have shown superior measurement qualities in comparison to traditional Likert-type response scales in previous studies. The present study expands the comparison of response scales to properties of Internet-based personality scales in a within-subjects design. A sample of 879 participants filled out an online questionnaire measuring Conscientiousness, Excitement Seeking, and Narcissism. The questionnaire contained all instruments in both answer scale versions in a counterbalanced design. Results show comparable reliabilities, means, and SDs for the VAS versions of the original scales, in comparison to Likert-type scales. To assess the validity of the measurements, age and gender were used as criteria, because all three constructs have shown non-zero correlations with age and gender in previous research. Both response scales showed a high overlap and the proposed relationships with age and gender. The associations were largely identical, with the exception of

an increase in explained variance when predicting age from the VAS version of Excitement Seeking ( $B_{10} = 1318.95$ ,  $\Delta R^2 = .025$ ). VASs showed similar properties to Likert-type response scales in most cases.

**Keywords** Visual analogue scale · VAS · Likert scale · Internet science · Measurement · Internet research · Online questionnaire · Response scale

---

The original version of this article was revised: In the Abstract, the penultimate sentence “The associations were largely identical, with the exception of an increase in explained variance when predicting age from the VAS version of Excitement Seeking ( $B_{10} = 125.1$ ,  $\Delta R^2 = .025$ .)” should be changed to “The associations were largely identical, with the exception of an increase in explained variance when predicting age from the VAS version of Excitement Seeking ( $B_{10} = 1318.95$ ,  $\Delta R^2 = .025$ .)” On page 4, Statistical analyses, at the end of the second paragraph, the sentence that reads “To avoid type I-error inflation due to multiple testing, the significance level was set to.” should be changed to “To avoid type I-error inflation due to multiple testing, the significance level was set to  $p < .01$ .”

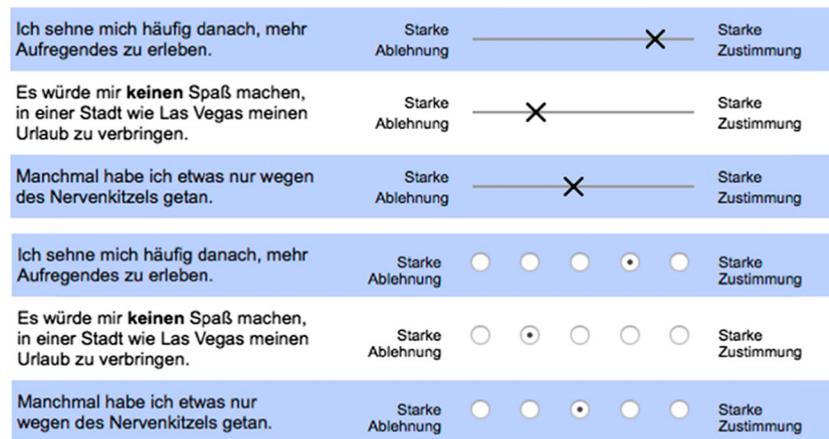
---

✉ Tim Kuhlmann  
tim.kuhlmann@uni-konstanz.de

<sup>1</sup> Research Methods, Assessment and iScience, Department of Psychology, University of Konstanz, Box 31, 78457 Konstanz, Germany

Conducting research via the Internet is considered a second revolution in behavioral research after the computer revolution in the late 1960s and early 1970s that brought about many advantages over widely used paper-and-pencil procedures (e.g., automated measurement, better precision). Internet-based questionnaires and tests are becoming more and more common in psychology and other social sciences (e.g., Reips, 2008; Reips & Birnbaum, 2011). When designing an online survey, the question of how to design items and which response options to present depends on a different set of rules than in offline surveys (e.g., Callegaro, Lozar Manfreda, & Vehovar, 2015; Hewson, Vogel, & Laurent, 2015). In measuring the degree of agreement with a statement, the response options are often presented in a Likert-type scale with a certain number of ordinal answer options. An alternative to Likert-type scales are visual analogue scales (VASs). These scales offer a number of advantages over Likert-type scales with regard to psychometric properties and are easy to implement online (Reips & Funke, 2008).

VASs are rating scales in a continuous graphical format. They were first described by Hayes and Paterson (1921). Instead of providing a discrete number of response options, the VAS provides participants with a straight line extending from one end of the scale (e.g., “Strongly Agree”) to the other (e.g., “Strongly Disagree”). Participants mark any point on this continuous line that corresponds to their subjective



**Fig. 1** Example of items in a VAS-format (**top**) and a Likert-type format (**bottom**)

agreement. In computer- and Internet-based assessments, each pixel in length of a VAS corresponds to a possible value (Reips & Funke, 2008). Examples of Likert and VAS response scale formats used in the present study are shown in Fig. 1. VASs are widely implemented in the medical sciences, especially when measuring changes in pain levels (e.g., Bijur, Silver, & Gallagher, 2001; Myles et al., 1999)

VASs allow respondents to communicate subjective values more exactly than radio button scales. Moreover, the format of VASs suggests to participants that their answer should be precise, because the number of response categories communicates how elaborated the expected answer should be. Correspondingly, a small number of response options implicitly conveys the message that roughly estimated answers are sufficient (Schwarz, 1999). Mean ratings tend to be equal in VASs and other response scales in paper-based studies (Cork et al., 2004) as well as in Internet-based studies (Funke & Reips, 2012; Kuhlmann, Reips, Wienert, & Lippke, 2016).

### Visual analogue scales and data quality

A theoretical advantage over discrete scales like radio button scales often used on the Internet is that answers are not restricted to a certain number of response options but very fine gradations can be measured. They avoid the systematic bias of values resulting from scale coarseness, a problem present when using items with distinct response categories that are meant to measure a continuous latent variable (Aguinis, Pierce, & Culpepper, 2009). In case the true value of the respondent is located between response options, Likert-type responses are potentially systematically biased upwards or downwards. This theoretical bias is reduced as the number of response categories is increased, though the psychological effect of presenting distinct response categories may still persist (Schwarz, 1999). In practice, most questionnaires using Likert-type scales

implement between four and seven response categories (Wakita, Ueshima, & Noguchi, 2012).

In paper-based studies on pain, Myles and colleagues (1999) and Myles and Urquhart (2005) found that data from VASs form a linear scale. Furthermore, participants were instructed to give a rating for feeling half or twice the amount of pain than their previous rating on a VAS that did not show the original score. Results from the comparison of these scores with the original ratings indicated equal intervals for these judgments. The authors conclude that "VAS scores can be treated as ratio data" (Myles et al., 1999). Reips and Funke (2008) also found that in Web questionnaire items, VASs fulfill the requirements of measurement on the level of an interval scale. Therefore, differences between ratings on VASs can be interpreted in a meaningful way and the prerequisites for many statistical procedures are met. Hayes, Allen, and Bennet (2013) describe the possibility of a general VAS to maximize the discrimination between neighboring categories and possibly produce less compressed answers at the extremes depending on the anchors used.

Given the cumulative evidence of better distributional properties and indications for a higher scale of measurement, the question remains whether VASs also offer more information about the assessed construct, i.e. whether responses gathered via VASs offer more valid information than those gathered via other response scales. This question was positively answered by Reips and Funke (2008) for percentages and further investigated by Funke and Reips (2012) for the construct *sense of style*. They compared the correlations between respondents' style regarding furnishing and clothing and found more positive correlations with VAS-type scales for construct related items and no correlations for construct-unrelated items when compared to their Likert counterparts. Thus, evidence showed that VASs are able to assess the domain general construct *sense of style* better.

The present study aims to investigate the additional information provided by responses on a VAS-type scale in comparison to responses to the same items on a Likert-type scale directly. In contrast to previous studies, which used a between-subjects design and compared correlations, the present study implements a within-subjects design allowing for a more detailed and consistent analysis of the effect of response scale type. To assess the validity of the information, a criterion with perfect reliability is the gold standard, because differences in the interrelation are directly related to the properties of the construct under investigation. The age and gender measures provide near-perfect reliability, which is why they were chosen for the present study.

### Relationship between personality characteristics and age and gender

Age and gender as variables are measured in most psychological studies along with other socio-demographic variables. They are often used to describe the sample and function as covariates or control variables in regression or ANCOVA analyses. In the present study, they are the main criterion variables in the analyses, because they possess the quality of a near-perfect criterion in terms of reliability. Several personality characteristics have shown a correlation with age and gender in previous research. The authors chose three well investigated and established personality constructs: Sensation Seeking, Conscientiousness, and Narcissism.

*Sensation Seeking* encompasses the tendency of individuals to seek out novel and exciting stimuli to reach their Optimal Level of Arousal (Zuckerman, Eysenck, & Eysenck, 1978). This has been shown to decline with increasing age after puberty, resulting in consistent negative correlations between age and Sensation Seeking (e.g., Steinberg et al., 2008; Zuckerman et al., 1978). Sensation Seeking has also been shown to be higher in males than in females (Steinberg et al., 2008; Zuckerman et al., 1978). The construct is measured as a subfacet of Extraversion in the Big Five personality model, labeled Excitement Seeking (Costa & McCrae, 1992).

The second personality trait investigated is the Big Five facet *Conscientiousness* (e.g., Buchanan, Johnson & Goldberg, 2005). This personality trait has also shown a relation to age in previous research. Srivastava et al. (2003) found a linear positive correlation throughout the whole lifespan, with some studies finding a trend towards a decline in late adulthood (Donnellan & Lucas, 2008). Srivastava et al. (2003) also found small gender differences for Conscientiousness, with women scoring slightly higher than men across the lifespan. This difference was also shown in a cross-cultural study by Donnellan and

Lucas (2008), though the difference ceased to show in late adulthood.

The third personality construct addressed in the present study is *Narcissism*, which has been shown to be negatively associated with age (e.g., Roberts et al. 2010; Twenge, Konrath, Foster, Campbell, & Bushman, 2008). There is an ongoing debate about whether these are developmental changes or cohort effects (Roberts et al. 2010; Twenge et al., 2008). In the context of the present study it is only relevant that the relation has been shown consistently. Consistent gender differences have also been shown for Narcissism, with men scoring higher than women on the whole measure (Grijalva et al., 2015). The gender difference is most pronounced in the facets Exploitative/Entitlement and Leadership/Authority. All of the aforementioned constructs have been investigated in numerous studies and measured by well validated and widely implemented scales.

### Research question and hypotheses

The main research question addressed in the current study is what we gain from implementing VASs as the response format, in comparison to Likert-type response scales. Specifically, the goal is to investigate whether VASs provide additional valid information when the information from the Likert-type scale is already accounted for. This extends the knowledge about better distributional properties and a better level of measurement for VASs to the area of validity and actual information contained in the measurement process. Our hypothesis proposes that personality instruments assessed via VAS-type response scales provide more information in comparison to the same instruments assessed via Likert-type response scales. This is expressed in additional explained variance when predicting the criteria age and gender. Furthermore, means and standard deviations (SDs) of the response-scale versions were compared as well as the intercorrelations between the constructs within and between both scale versions.

### Method

#### Items in Likert or VAS format

The experimental manipulation in the present study was the implementation of response scales in two different formats for the measured constructs. Scales were presented in a Likert format with five radio buttons or in a VAS format using the same anchors and same item wording. Both response scales covered the same width on-screen. Examples of three Excitement Seeking items in both response scale formats are shown in Fig. 1.

## Procedure

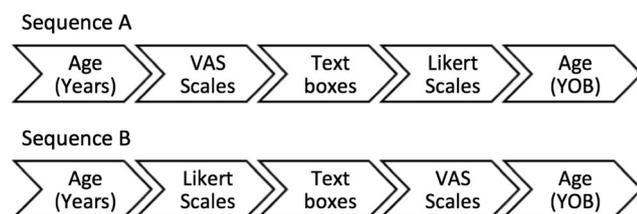
Recruitment for the Internet-based questionnaire was carried out during a seminar. Participating students received course credit and were instructed to acquire at least 20 participants from the general population. Due to the research design requiring variance in the age of participants, they were instructed to recruit participants of varying ages to counter the potential problem of range restriction in this dependent measure. The questionnaire contained the seriousness check (Aust, Diedenhofen, Ullrich, & Musch, 2013; Reips, 2002, 2009) at the beginning. Participants then indicated their age and gender, before the personality scales were presented. No other sociodemographic variables were assessed. Each participant responded twice to each scale, once in a Likert and once in a VAS version. The order of versions in the within-subject design was counterbalanced. Before answering the scales for a second time, an unrelated scale was presented. This scale used text fields as the response format, avoiding a possible confounding factor in the counterbalanced design. To evaluate the reliability of the assessment of age, participants were asked at the end of the questionnaire to provide their year of birth. An overview of the sequence of scales for the two counterbalanced versions is shown in Fig. 2. Participants on average took 9.9 min ( $SD = 6.7$ ) to complete the survey. No remuneration was provided.

## Measures

The questionnaire contained scales of well-established personality instruments. These included the Excitement Seeking subscale of the Big Five facet Extraversion from the NEO-PI-R (Ostendorf & Angleitner, 2004). The subscale consists of eight items. Furthermore, Conscientiousness was measured using the 12-item scale of the NEO-FFI (Borkenau & Ostendorf, 1993). The six-item short version of the NARQ (Back et al., 2013) was implemented to measure Narcissism.

## Sample characteristics

The sample consisted of 879 participants of which 75 indicated in the seriousness check that they did not want to participate seriously. These cases were not included with the



**Fig. 2** Flow chart of the counterbalanced progression through the questionnaire

analyses. Age ranged from 14–82 years ( $M = 33$ ,  $SD=14.7$ ). Female participants comprised 61.8% of the sample. The only exclusion criterion applied was that participants had to be proficient in German in order to ascertain understanding of the items. The study was conducted in accordance with the principles laid down in the Declaration of Helsinki (World Medical Association, 2013) and with the institutional guidelines of the Department of Psychology, University of Konstanz.

## Statistical analyses

To evaluate the Likert and VAS version of each scale, their correlation with age and gender was computed to ascertain the relationships established by previous studies. To test for incremental variance in the association that the VAS versions of the scale explain above the Likert version, hierarchical linear Bayesian Regression analyses with age as the dependent variable were performed. We compared the model that includes all Likert-type response scales as predictors with the model that additionally includes all VAS-type scales. The comparison was evaluated by the Bayes factor comparing both models. To interpret the results, we used the evidence categorization proposed by Lee and Wagenmakers (2013). A Bayes factor of one signifies no preference for either the null or the alternative model. Bayes factors different from one indicate evidence towards one of the two compared models, with values higher than ten and lower than one-tenth being classified as strong evidence. The analyses were carried out in R using the BayesFactor package (Morey & Rouder, 2015) with the default priors. Gender was predicted via Bayesian logistic regression implementing estimation procedures proposed by Kruschke (2015). As the method proposed by Kruschke (2015) does not allow for direct model comparison, the regression coefficients of individual logistic regressions were compared using the highest density intervals (HDIs) for the estimates.

In case of the model with VAS-type scales showing an improvement, separate hierarchical regressions were performed with either age or gender as the dependent variable to investigate the effect in detail. As collinearity of the predictors inflates the standard errors and therefore prevents reliable parameter estimation (Gujarati, 2003), the values of the VAS were regressed on the corresponding Likert-scale. Instead of entering the VAS values in the second step, only the residuals of the VAS were entered in the second step. This represents the amount of information contained in the VAS measurement that is not already explained by the Likert scale and allows investigation of the direction of the effect. To avoid type I-error inflation due to multiple testing, the significance level was set to  $p < .01$ . The dataset is available at <https://osf.io/gvqjs/>.

## Results

The results of the analyses are presented below. These include the consistency of the age measure and the reliabilities for all scales in both response scale versions. Afterwards the results regarding intercorrelations, distributional properties, and regression analyses are presented. These analyses were carried out separately for the counterbalanced versions of the questionnaire. No differences emerged. Therefore we report results for the data collapsed over both versions.

### Consistency of the age measure

To assess the accuracy of self-reported age, the year of birth was recoded into age by subtracting it from the year of the assessment. The age was considered consistent if the two age measures matched exactly or if the measure derived from the year of birth was 1 year older. This accounts for the fact that a participant's birthday might not have taken place before the date of the study. Both ages matched in 98.3% of cases, increasing to 99.7% in case a difference of 1 year was coded as marginal. The two cases with an age mismatch according to this marginality criterion were excluded from further analyses.

### Psychometric properties of the scales

Internal consistencies for all three personality constructs differentiated by response scale are shown in Table 1. Internal consistencies are similar for both types of response scales.

### Correlations of the constructs measured via different response scales

The intercorrelations of all personality measures for both response scale formats are shown in Table 2. Measurement of the same construct via different response scales shows a high overlap. Depending on the construct measured, 81.0–86.5% of the variance was shared by the scales in different response formats.

The correlations between the three personality scales are also shown. The pattern of associations is nearly identical between the different response scale versions with none of the correlations significantly differing regardless of the response scale version used for the comparison (all  $p$ 's > .1).

**Table 1** Cronbach's alpha of the personality scales by response format

	VAS	Likert
Excitement Seeking	.68	.67
Conscientiousness	.84	.82
Narcissism	.78	.75

**Table 2** Intercorrelations of personality measures across scale versions

		VAS			Likert		
		ES	C	Nar	ES	C	Nar
VAS	Excitement Seeking (ES)	-	-.11	.37	.94	-.11	.36
	Conscientiousness (C)		-	-.13	-.10	.90	-.12
	Narcissism (Nar)			-	.35	-.13	.90
Likert	Excitement Seeking (ES)				-	-.11	.35
	Conscientiousness (C)					-	-.14
	Narcissism (Nar)						-

Note. N = 584–595

### Means and standard deviations

Means and SDs of constructs assessed via VAS-type scales have been compared to the same constructs assessed via Likert-type scales in between-subjects designs in previous studies (Funke & Reips, 2012; Kuhlmann, Reips, Wienert, & Lippke, 2016). The means and SDs of the constructs investigated are depicted for both response-scale formats in Table 3. The present study investigated the influence of the scale version on the mean score via a repeated measures ANOVA design. Responses of the Likert-type response scale were recoded into corresponding VAS-values, 1–101, via linear transformation. The repeated measures factor was the response scale version of the same construct, VAS versus Likert. The dependent variables were the three personality constructs, Excitement Seeking, Narcissism, and Conscientiousness. There was no significant main effect of response scale,  $F(1, 579) = 1.99$ ,  $p = .16$ . The interaction between the response scale and personality construct was also not significant,  $F(2, 1158) = .43$ ,  $p = .65$ . Given the high power, this indicates no or only a very small influence of response scale format on the mean score of the constructs. A Bayesian repeated measures ANOVA with the same design favored the null hypothesis of no difference between the response scale versions with a Bayes Factor of  $B_{01} = 22.9$ .

**Table 3** Means and standard deviations (SDs) of the three personality scales in the two response formats

Construct (Response scale)	Mean	SD
Excitement Seeking (VAS)	49.0	17.8
Excitement Seeking (Likert)	49.3	18.2
Conscientiousness (VAS)	72.9	13.8
Conscientiousness (Likert)	72.9	13.7
Narcissism (VAS)	29.5	16.7
Narcissism (Likert)	29.9	17.1

Notes. Response range from 1–101, Likert scale values recoded via linear transformation

**Table 4** Correlations with age for VAS and Likert versions of the personality scales

	VAS	Likert
Excitement Seeking	-.53	-.51
Conscientiousness	.15	.17
Narcissism	-.25	-.26

Note. All coefficients are significant with  $p < .01$

indicating strong evidence for no effect of scale version (Lee & Wagenmakers, 2013). The effect size estimates for the differences between response scale versions were low, from  $d = .02$  to  $d = .05$ .

SDs of the different response scale formats were compared using Levene's test of homogeneity of variance. Contrary to previous studies, no difference in the magnitude of SDs emerged for either personality scale (all  $p$ 's  $> .1$ ).

### Regression of age on scales

In order to compare the explained variance in the dependent variable of age, we first examined the correlations between the personality scales and the age measure to ascertain that the proposed relationship from previous studies was also present in our sample. This was true in the case of all three personality constructs for both versions of the scale. The correlations for all constructs in both response scale versions are shown in Table 4.

To examine the additional information present in the response on the VAS in comparison to the Likert scale, a Bayesian hierarchical regression with all six measures, three personality measures in both response scale versions, was performed. The null model included the three personality constructs measured via a Likert-type scale; results are shown in Table 5. The model that includes all VAS-type scales is favored over the null model by a Bayes factor of 29.74, providing strong evidence for the inclusion of the VASs. This advantage stems only from the Excitement Seeking VAS-scale, as for both the Conscientiousness and Narcissism scales the null model including only Likert-type scales is favored,  $B_{10} = .13$

**Table 5** Hierarchical Bayesian regression of age on Likert- and VAS-scale versions

Models	BF <sub>10</sub>
Null model (including all Likert scales)	1.000
Adding all VASs simultaneously	29.74
Adding Excitement Seeking VAS	1318.95
Adding Conscientiousness VAS	0.13
Adding Narcissism VAS	0.11

Note.  $BF_{10}$  = Bayes factor for the model including the variable(s) over the null model

**Table 6** Hierarchical regression of age on Excitement Seeking

	Excitement Seeking	
	Beta	$\Delta R^2$
Step 1		
Likert scale	-.51***	-
Step 2		.025***
Likert scale	-.51***	
VAS residual	-.16***	

Note. \*\*\* =  $p < .001$

and .11, respectively. The model including the VAS Excitement Seeking scale is favored by a Bayes factor of  $B_{10} = 1318.95$ , considered as extreme evidence for inclusion.

When entering the VAS-type scales first and the Likert-type scales in the second step, the null model including only VASs was preferred by a Bayes factor of  $B_{01} = 31.35$  (very strong evidence).

To further examine the additional informational value the Excitement Seeking VAS-scale version provides over the Likert-type scale, the values of the VAS were linearly regressed on the scale values of the corresponding Likert scale. The predicted VAS value was then subtracted from the actual value on the VAS for each participant. This resulted in a residual VAS term of the participant, which could not be explained by the Likert value of the scale. In a hierarchical regression analysis with age as the dependent variable, we then entered the Likert scale value of the personality construct in the first step. In the second step the residual VAS term was entered to examine whether additional variance was explained. The resulting regressions are shown in Table 6.

The residuals accounted for about 2.5% additional variance explained for Excitement Seeking. The direction of the effect for the residuals is the same as that of the main scale, indicating valid additional variance.

### Regression of gender on scales

The analyses regarding gender were carried out analogous to the ones with age as the dependent variable. Because gender is coded as dichotomous, logistic regressions were performed. First, the point-biserial correlations between the personality constructs and gender were computed. These are shown in Table 7.

**Table 7** Point-biserial correlations with gender for VAS and Likert versions of the personality scales

	VAS	Likert
Excitement Seeking	.08	.09*
Conscientiousness	.00	.00
Narcissism	.25***	.23***

Notes \* =  $p < .05$ ; \*\*\* =  $p < .001$

**Table 8** Logistic Bayesian regression of gender on Excitement Seeking and Narcissism for both response scales

Predictor	Estimate ( $\beta$ )	95% HDI	99% HDI
Excitement Seeking VAS	.010	[.000, .019]	[-.003, .022]
Excitement Seeking Likert	.011	[.002, .020]	[-.001, .024]
Narcissism VAS	.031	[.020, .041]	[.017, .046]
Narcissism Likert	.028	[.018, .038]	[.015, .042]

Note. HDI highest density interval

In our sample gender was not associated with levels of Conscientiousness and only a small association with Excitement Seeking was measured. The regression analyses were therefore only carried out with Narcissism and Excitement Seeking as predictors. The results for the Bayesian logistic regression with the Likert and VAS values entered are shown in Table 8.

The logistic regressions for Excitement Seeking did not support either of the scale versions as a predictor of gender. The 99% and 95% HDIs for each scale version estimate do include the estimate for the respective other scale, however. Narcissism was shown to be a significant predictor of gender for both scale versions, with the HDIs also indicating no difference between the estimates for either scale.

## Discussion

The present study aimed at investigating the properties of VAS-type response scales in a within-subjects design, comparing them to the more traditional Likert-type scales. The two versions showed considerable overlap as indicated by the high correlations between both versions of the scales. Internal reliabilities, means, and SDs did not differ between the response-scale versions. The intercorrelations between the three personality measures were also identical independent of the response scale used for measurement. All scales showed the expected relationship with the age measure. Narcissism and Excitement Seeking but not Conscientiousness were associated with the second criterion, gender.

The informational value of both scale versions emerged as identical, with the exception of Excitement Seeking. In this case the prediction of age was improved for the VAS version of the scale as shown by traditional as well as Bayesian regression models. When predicting gender, only Narcissism was a significant predictor with equal estimates for the Likert-type and VAS versions. The results of the present study indicate equivalence in measurement and associations between Likert-type scales and VASs, the only exception being the benefit of the VAS Excitement Seeking scale when predicting age. The effect size was small, though, with only 2.5% incremental variance explained out of a total of 28.6% explained variance. A possible

explanation for this finding is that it emerged for the strongest original association, as age and Excitement Seeking had the highest correlation in the present sample,  $r = -.53$  to  $-.51$ . Further research is necessary to evaluate the robustness and possible explanations for this effect.

The analyses in the present study use age and gender as reliable criteria. The reliability of the age measure was confirmed by a very high consistency of entries at the beginning and end of the questionnaire. Differences in the relationship between the personality scales and age and gender are only dependent on the information provided by the measurement of personality.

Our results indicate that the more fine-grained responses that are possible via VASs (Funke & Reips, 2012) did not lead to a better measurement in most cases. They did not increase the error variance in responses either, as shown by very similar associations within the scales as well as with the criteria age and gender. The equality of means and SDs also suggests similar distributional properties of both response scale versions. The finding of equal SDs contrasts with previous research that found lower SDs when implementing VASs (Funke & Reips, 2012; Kuhlmann, Reips, Wienert, & Lippke, 2016). A possible explanation for this finding is the difference between the assessed constructs. The present study measured personality characteristics, whereas the previous studies investigated health-related variables and semantic differentials. Another possible explanation is the difference in design. The present study investigated both response scales in a within-subjects design, previous studies implemented between-subjects designs.

The present study provides evidence for equal measurement when implementing VASs in comparison to Likert-type scales in a within-subjects design. Further research should be conducted to investigate other variables and ascertain the validity of VASs. Examining the impact of the length of the scale, including longer questionnaires, is also a possible direction for future studies. It should be noted, however, that the results with regard to external criteria only include age and gender. This renders the interpretation of the results more difficult and presents several possible explanations.

## Limitations

The present study used a voluntary recruitment strategy through participants of a seminar. Since this is not a random sampling method, the generalizability of the results is limited. Recruiters were instructed not to disclose the purpose of the study, but naivety of study participants could not be guaranteed. On the other hand, the present sample does extend beyond the usual limits of a student sample because a wider age range was recruited.

A second limitation is the scope of the present study. We only investigated three personality scales and their relationship with gender and age. Construct validity remains to be investigated in future research. Age and gender were chosen as external criteria

in the current study as they provide near-perfect reliability. Other criteria were not investigated in the present study, though evidence for informational advantages when predicting age and gender could very well translate to other criteria.

A third limitation is that only linear trends were investigated in the present study. The relationship between the response scales and age were examined via correlations and the GLM. This represents the most common analyses performed, but does not take into account possible nonlinear relationships between responses on a Likert scale and responses on a VAS-type scale. There is no theoretical reason to assume a meaningful nonlinear trend; nonetheless our analyses are not able to account for it in case it exists.

## Conclusion

The present study is the first to investigate Likert-type and VAS-type scales in a within-subjects design, thus allowing for a direct comparison. The results indicate that VAS- and Likert-response scales provide identical information and have identical distributional qualities. As in most previous research, results indicate no disadvantages to implementing VASs. Further research should examine other personality measurements and validation criteria to examine possible moderator effects.

## References

- Aguinis, H., Pierce, C. A., & Culpepper, S. A. (2009). Scale coarseness as a methodological artifact: Correcting correlation coefficients attenuated from using coarse scales. *Organizational Research Methods, 12*, 623–652. doi:10.1177/1094428108318065
- Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods, 45*, 527–535. doi:10.3758/s13428-012-0265-2
- Back, M. D., Küfner, A. C. P., Dufner, M., Gerlach, T. M., Rauthmann, J. F., & Denissen, J. J. A. (2013). Narcissistic admiration and rivalry: Disentangling the bright and dark sides of narcissism. *Journal of Personality and Social Psychology, 105*, 1013–1037. doi:10.1037/a0034431
- Bijur, P. E., Silver, W., & Gallagher, E. J. (2001). Reliability of the visual analog scale for measurement of acute pain. *Academic Emergency Medicine, 8*, 1153–1157. doi:10.1111/j.1553-2712.2001.tb01132.x
- Borkenau, P., & Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar (NEO-FFI); nach Costa und McCrae [NEO-five-factor inventory (NEO-FFI); based on Costa and McCrae]*. Göttingen: Hogrefe.
- Buchanan, T., Johnson, J. A., & Goldberg, L. R. (2005). Implementing a five-factor personality inventory for use on the Internet. *European Journal of Psychological Assessment, 21*, 115–127. doi:10.1027/1015-5759.18.1.115
- Callegaro, M., Manfreda, K. L., & Vehovar, V. (2015). *Web survey methodology*. Los Angeles, CA: Sage.
- Cork, R. C., Isaac, I., Elsharydah, A., Saleemi, S., Zavisca, F., & Alexander, L. (2004). A comparison of the verbal rating scale and the visual analog scale for pain assessment. *The Internet Journal of Anesthesiology, 8*, 23–38.
- Costa, P. T., & McCrae, R. R. (1992). *Neo PI-R professional manual*. Odessa, FL: Psychological Assessment Resources.
- Donnellan, M. B., & Lucas, R. E. (2008). Age differences in the big five across the life span: Evidence from two national samples. *Psychology and Aging, 23*, 558–566. doi:10.1037/a0012897
- Funke, F., & Reips, U.-D. (2012). Why semantic differentials in web-based research should be made from visual analogue scales and not from 5-point scales. *Field Methods, 24*, 310–327. doi:10.1177/1525822X12444061
- Grijalva, E., Newman, D. A., Tay, L., Donnellan, M. B., Harms, P. D., Robins, R. W., & Yan, T. (2015). Gender differences in narcissism: A meta-analytic review. *Psychological Bulletin, 141*, 261–310. doi:10.1037/a0038231
- Gujarati, D. N. (2003). *Multicollinearity: What happens if the regressors are correlated?* *Basic Econometrics, 341–386*. Boston, MA, US: McGraw-Hill.
- Hayes, M. H., & Patterson, D. G. (1921). Experimental development of the graphic rating method. *Psychological Bulletin, 18*, 98–99.
- Hayes, J. E., Allen, A. L., & Bennett, S. M. (2013). Direct comparison of the generalized visual analog scale (gVAS) and general labeled magnitude scale (gLMS). *Food Quality and Preference, 28*, 36–44. doi:10.1016/j.foodqual.2012.07.012
- Hewson, C., Vogel, C., & Laurent, D. (2015). *Internet research methods*. London: Sage.
- Kruschke, J. K. (2015). *Doing Bayesian Data Analysis, Second Edition: A Tutorial with R, JAGS, and Stan*. London: Academic Press/Elsevier.
- Kuhlmann, T., Reips, U.-D., Wienert, J., & Lippke, S. (2016). Using Visual Analogue Scales in eHealth: Non-response effects in a lifestyle intervention. *Journal of Medical Internet Research*.
- Lee, M. D., & Wagenmakers, E. J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge: Cambridge University Press.
- Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of Bayes Factors for Common Designs*. R package version 0.9.12-2.
- Myles, P. S., & Urquhart, N. (2005). The linearity of the visual analogue scale in patients with severe acute pain. *Anaesthesia and Intensive Care, 33*, 54–58.
- Myles, P. S., Troedel, S., Boquest, M., & Reeves, M. (1999). The pain visual analog scale: Is it linear or nonlinear? *Anesthesia & Analgesia, 89*, 1517–1520.
- Ostendorf, F., & Angleitner, A. (2004). *NEO-Persönlichkeitsinventar nach Costa und McCrae, Revidierte Fassung (NEO-PI-R) [NEO-personality inventory based on Costa und McCrae, revised version (NEO-PI-R)]*. Göttingen: Hogrefe.
- Reips, U.-D. (2002). Standards for Internet-based experimenting. *Experimental Psychology, 49*, 243–256.
- Reips, U.-D. (2008). How Internet-mediated research changes science. In A. Barak (Ed.), *Psychological aspects of cyberspace: Theory, research, applications* (pp. 268–294). Cambridge: Cambridge University Press.
- Reips, U.-D. (2009). Internet experiments: Methods, guidelines, metadata. *Human Vision and Electronic Imaging XIV, Proceedings of SPIE, 7240*, 724008.
- Reips, U.-D., & Birnbaum, M. H. (2011). Behavioral research and data collection via the Internet. In K.-L. Vu & R. W. Proctor (Eds.), *The handbook of human factors in web design* (2nd ed., pp. 563–585). Mahwah, New Jersey: Erlbaum.
- Reips, U.-D., & Funke, F. (2008). Interval-level measurement with visual analogue scales in Internet-based research: VAS generator. *Behavior Research Methods, 40*, 699–704. doi:10.3758/BRM.40.3.699
- Roberts, B. W., Edmonds, G., & Grijalva, E. (2010). It is developmental me, not Generation Me developmental changes are more important than generational changes in Narcissism—Commentary on Trzesniewski & Donnellan (2010). *Perspectives on Psychological Science, 5*, 97–102. doi:10.1177/1745691609357019

- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*, 93–105.
- Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology*, *84*, 1041–1053. doi:10.1037/0022-3514.84.5.1041
- Steinberg, L., Albert, D., Cauffman, E., Banich, M., Graham, S., & Woolard, J. (2008). Age differences in sensation seeking and impulsivity as indexed by behavior and self-report: Evidence for a dual systems model. *Developmental Psychology*, *44*, 1764–1778. doi:10.1037/a0012955
- Twenge, J. M., Konrath, S., Foster, J. D., Keith Campbell, W., & Bushman, B. J. (2008). Egos inflating over time: A cross-temporal meta-analysis of the narcissistic personality inventory. *Journal of Personality*, *76*, 875–902. doi:10.1111/j.1467-6494.2008.00507.x
- Wakita, T., Ueshima, N., & Noguchi, H. (2012). Psychological distance between categories in the likert scale comparing different numbers of options. *Educational and Psychological Measurement*, *72*, 533–546. doi:10.1177/0013164411431162
- World Medical Association. (2013). *WMA declaration of Helsinki: Ethical principles for medical research involving human subjects*. Retrieved from <http://www.wma.net/en/30publications/10policies/b3/>
- Zuckerman, M., Eysenck, S. B., & Eysenck, H. J. (1978). Sensation seeking in England and America: Cross-cultural, age, and sex comparisons. *Journal of Consulting and Clinical Psychology*, *46*, 139–149. doi:10.1037/0022-006X.46.1.139