Do visual analogue scales perform better than Likert-type scales?
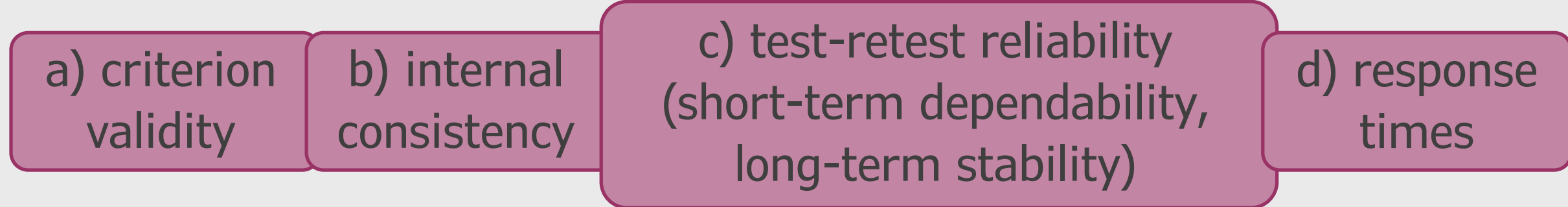
One might consider a Visual Analogue Scale (VAS) as a better option for measuring continuous latent traits compared to a Likert Scale (LS). Since the VAS is essentially a LS with an infinite number of points, it offers a more nuanced way to respond, increases systematic variance, and better approximates the continuous nature of data (Chyung et al., 2018). Hence, some authors (e.g., Funke & Reips, 2012; Haslbeck et al., 2024; Kuhlmann et al., 2016) recommend prioritizing VAS over LS. However, our within-subject experiment indicates that...

…the **Visual Analogue Scale does not outperform the Likert Scale** in criterion validity, internal consistency, stability, or dependability of test scores. It only **takes more time**.
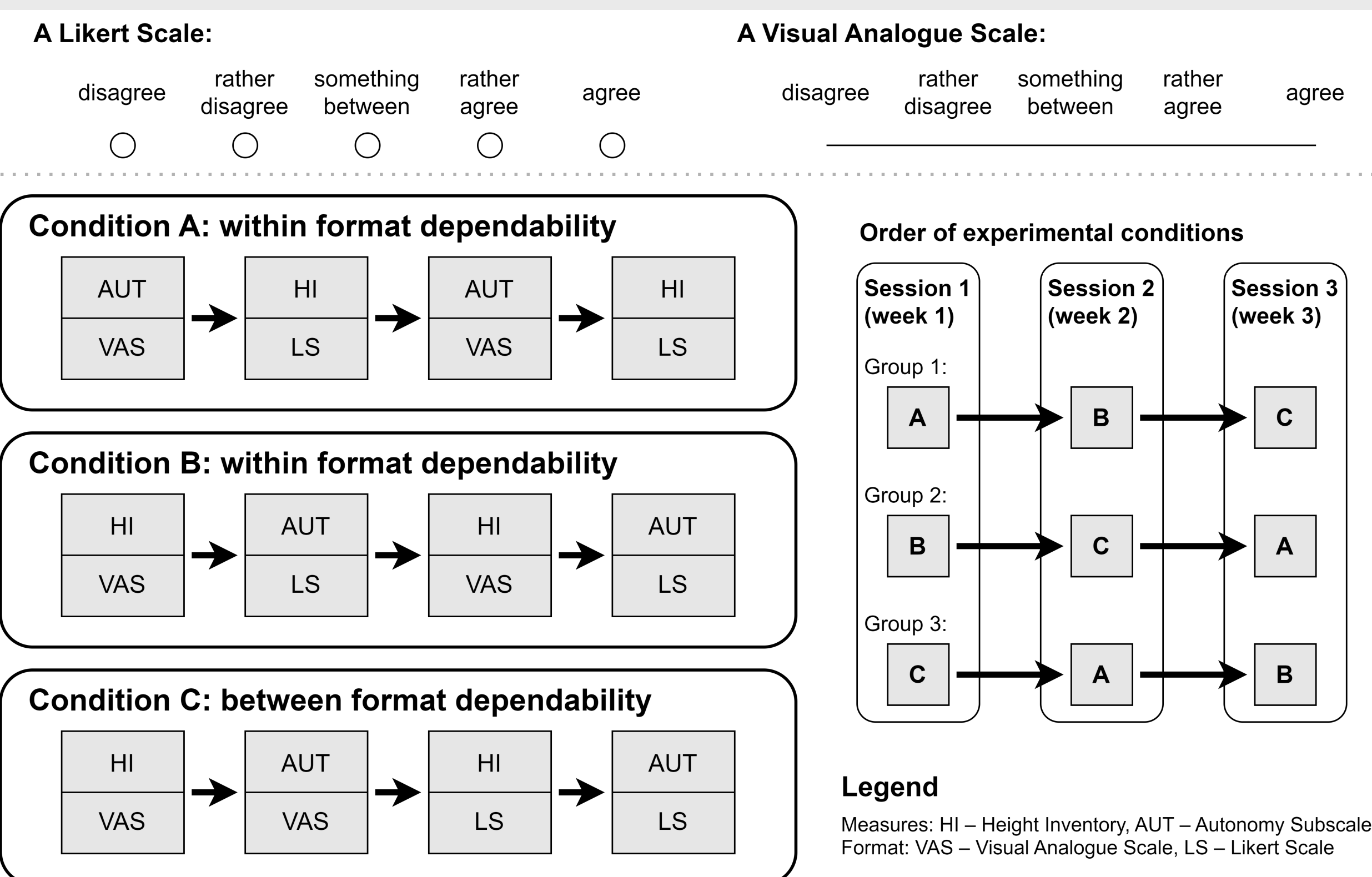
INTRODUCTION

- contradictory findings on internal consistency, descriptive statistics, and respondents' preferences (e.g., Hilbert et al., 2016; Kuhlmann et al., 2017; Lewis & Erdinç, 2017; Preston & Colman, 2000; Simms et al., 2019; Voutilainen et al., 2016)
- lacking evidence for the criterion validity and temporal stability
- this study aims to compare VAS and LS in terms of:

  - a) criterion validity
  - b) internal consistency
  - c) test-retest reliability (short-term dependability, long-term stability)
  - d) response times

METHOD

- *N* = 980, online within-subject experiment
- measures: **Height Inventory** (Rečka, 2018; Tancoš, 2019), **Autonomy Subscale** from the Basic Needs Satisfaction in General Scale (Gagné, 2003) + **additional data** (e.g., self-reported height)
- response formats: a **5-point LS**, a **VAS with verbal anchors**

A Likert Scale:

disagree rather disagree something between rather agree agree

A Visual Analogue Scale:

disagree rather disagree something between rather agree agree

Condition A: within format dependability

AUT / VAS → HI / LS → AUT / VAS → HI / LS

Condition B: within format dependability

HI / VAS → AUT / LS → HI / VAS → AUT / LS

Condition C: between format dependability

HI / VAS → AUT / VAS → HI / LS → AUT / LS

Order of experimental conditions

Session 1 (week 1)   Session 2 (week 2)   Session 3 (week 3)

Group 1:   A → B → C

Group 2:   B → C → A

Group 3:   C → A → B

Legend

Measures: HI – Height Inventory, AUT – Autonomy Subscale
Format: VAS – Visual Analogue Scale, LS – Likert Scale

ANALYSES & RESULTS

- a) criterion validity from the realistic perspective (path analyses on the HI sum score ~ self-reported height) – comparable results: for LS: r_{men} = .91, r_{women} = .90; for VAS: the same

- b, c) internal consistency (Cronbach's α, ordinal α by Zumbo et al., 2007), test-retest reliabilities (pairwise correlation matrices) – comparable results

		condition A	condition A	condition B	condition B	condition C	condition C
		LS	LS	VAS	VAS	VAS	LS
condition A	LS	HI $\alpha_♂$ = .91 $\alpha_♀$ = .92	DEP_W r = .94	STB_B r = .87	STB_B r = .87	STB_B r = .87	STB_W r = .87
condition A	LS	DEP_W $r_♂$ = .99 $r_♀$ = .98	HI $\alpha_♂$ = .92 $\alpha_♀$ = .94	STB_B r = .87	STB_B r = .87	STB_B r = .87	STB_W r = .87
condition B	VAS	STB_B $r_♂$ = .97 $r_♀$ = .97	STB_B $r_♂$ = .97 $r_♀$ = .97	HI $\alpha_♂$ = .91 $\alpha_♀$ = .92	DEP_W r = .95	STB_W r = .88	STB_B r = .87
condition B	VAS	STB_B $r_♂$ = .97 $r_♀$ = .97	STB_B $r_♂$ = .97 $r_♀$ = .97	DEP_W $r_♂$ = .98 $r_♀$ = .99	HI $\alpha_♂$ = .91 $\alpha_♀$ = .92	STB_W r = .88	STB_B r = .87
condition C	VAS	STB_B $r_♂$ = .97 $r_♀$ = .97	STB_B $r_♂$ = .97 $r_♀$ = .97	STB_W $r_♂$ = .97 $r_♀$ = .97	STB_W $r_♂$ = .97 $r_♀$ = .97	AUT α = .84	DEP_B r = .94
condition C	LS	STB_W $r_♂$ = .97 $r_♀$ = .97	STB_W $r_♂$ = .97 $r_♀$ = .97	STB_B $r_♂$ = .97 $r_♀$ = .97	STB_B $r_♂$ = .97 $r_♀$ = .97	DEP_B $r_♂$ = .98 $r_♀$ = .98	AUT α = .84 α_o = .88

Type of reliability
DEP – dependability (blue)
STB – stability (green)
α – internal consistency (red),
    α – Cronbach α, α_o – ordinal α

Measure
HI – Height Inventory (below the bold diagonal)
AUT – Autonomy Subscale (above it)

Response format
LS – Likert Scale
VAS – Visual Analogue Scale

Same or different format
W – within-response format test-retest
    (LS – light color, VAS – dark color)
B – between-response format test-retest
    (medium dark color)

Sample
♂ – a subset of men
♀ – a subset of women

Note. Where the coefficient has been estimated multiple times due to the design of the experiment (i.e., for internal consistency), we provide the average of these estimates.

- d) response times (log-linear mixed-effect models) – VAS took 16% longer than LS (p < .001)

DISCUSSION & CONCLUSION

- Both LS and VAS proved to be realistically valid and reliable with only negligible differences. The only exception was the response time (in favor of the LS). Similarly to Simms et al. (2019), we conclude that the VAS has no advantages over the LS. Due to greater time demands of VAS, **we recommend using LS.**

REFERENCES
Chyung, S. Y., Swanson, I., Roberts, K., & Hankinson, A. (2018). Evidence-based survey design: The use of continuous rating scales in surveys. *Performance Improvement, 57*(5), 38–48. https://doi.org/10.1002/pfi.21763 • 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*(3), 310–327. https://doi.org/10.1177/1525822X12444061 • Gagné, M. (2003). The role of autonomy support and autonomy orientation in prosocial behavior engagement. *Motivation and Emotion, 27*(3), 199–223. https://doi.org/10.1023/A:1025007614869 • Haslbeck, J. M. B., Jover Martínez, A., Roefs, A. J., Fried, E. I., Lemmens, L. H. J. M., Groot, E. L., & Edelsbrunner, P. A. (2024). Comparing Likert and visual analogue scales in ecological momentary assessment. PsyArXiv. https://doi.org/10.31234/osf.io/yt8xw • Hilbert, S., Küchenhoff, H., Sarubin, N., Nakagawa, T. T., & Bühner, M. (2016). The influence of the response format in a personality questionnaire: An analysis of a dichotomous, a Likert-type, and a visual analogue scale. *Testing, Psychometrics, Methodology in Applied Psychology, 23*(1), 3–24. https://doi.org/10.4473/TPM23.1.1 • Kuhlmann, T., Dantlgraber, M., & Reips, U.-D. (2017). Investigating measurement equivalence of visual analogue scales and Likert-type scales in Internet-based personality questionnaires. *Behavior Research Methods, 49*(6), 2173–2181. https://doi.org/10.3758/s13428-016-0850-x • 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, 18*(6), e126. https://doi.org/10.2196/jmir.527 • Lewis, J. R., & Erdinç, O. (2017). User experience categories in rating scales with 7, 11, or 101 points: Does it matter? *Journal of Usability Studies, 12*(2), 73–91. • Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica, 104*(1), 1–15. https://doi.org/10.1016/S0001-6918(99)00050-5 • Rečka, K. (2018). *Dotazník výšky a váhy* [Magisterská diplomová práce, Masarykova univerzita]. Institucionální repozitář MU. https://is.muni.cz/th/ug7c2/ • Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychological Assessment, 31*(4), 557–566. https://doi.org/10.1037/pas0000648 • Tancoš, M. (2019). *Vliv verbálních kotev Likertovy škály na psychometrické charakteristiky dotazníků* [Bakalářská diplomová práce, Masarykova univerzita]. Institucionální repozitář MU. https://is.muni.cz/th/uk8cb/ • Voutilainen, A., Pitkäaho, T., Kvist, T., & Vehviläinen-Julkunen, K. (2016). How to ask about patient satisfaction? The visual analogue scale is less vulnerable to confounding factors and ceiling effect than a symmetric Likert scale. *Journal of Advanced Nursing, 72*(4), 946–957. https://doi.org/10.1111/jan.12875 • Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods, 6*(1), 21–29. https://doi.org/10.22237/jmasm/1177992180

Eva Šragová, Hynek Cígler, David Elek, Petra Hubatka, Gabriela Kalistová | sragova@mail.muni.cz | Masaryk University, Brno, Czechia