Contextual variation in cognitive performance of older adults: Demonstration of an age-of-examiner effect

Olivier Desrichard, Neele Heiser, Olivier Renaud, Sascha Zuber, Michel Oris & Matthias Kliegel

Routledge
Taylor & Francis Group

🔓 OPEN ACCESS   🔴 Check for updates

Contextual variation in cognitive performance of older adults: Demonstration of an age-of-examiner effect

Olivier Desrichard[a,b,c], Neele Heiser[a], Olivier Renaud[a,b], Sascha Zuber[a,b,c], Michel Oris[b,c,d] and Matthias Kliegel[a,b,c] 🆔

[a]Department of Psychology, University of Geneva, Geneva, Switzerland; [b]Swiss National Center of Competences in Research LIVES–Overcoming Vulnerability: Life Course Perspectives, Lausanne and Geneva, Switzerland; [c]Center for the Interdisciplinary Study of Gerontology and Vulnerability, University of Geneva, Geneva, Switzerland; [d]Faculty of Social Sciences, University of Geneva, Geneva, Switzerland

ABSTRACT

Objective: Consistent with research on stereotype threat, when examiners' characteristics make a stereotype of the participant group salient, it can hamper participants' performance. We hypothesized that younger examiners represent a subtle element activating age stereotypes, leading older people to perform worse as examiners' age decreases. **Method:** We analyzed data from the Survey of Health, Ageing, and Retirement in Europe (SHARE; $N_{Participants}$ = 32768) and Vivre-Leben-Vivere studies (VLV, $N_{participants}$ = 960), wherein older people were tested at home by examiners of different ages on eight cognitive tasks. **Results:** Our results indicate that participants' performance on five tasks was positively linked to examiners' age, showing that the older the examiner, the better the participants' performance. **Conclusions:** These findings could have implications for the current assessment of memory performance among older adults.

Measuring the cognitive abilities of older adults is a common practice for diagnosing neurodegenerative diseases and studying aging. However, variations in older adults cognitive test performance due to variations in the testing environment have been increasingly documented (Fresson et al., 2017; Schlemmer & Desrichard, 2018; Sindi et al., 2013; VanLandingham et al., 2022). A source of variation is a situation's potential to make an age-based stereotype salient. As proposed by Levy (Barber et al., 2020; Levy, 2003, 2009), from an early age, people are exposed to positive and negative beliefs about aging. In old age, these typically internalized beliefs become self-relevant and the basis for self-perception (Levy, 2009). Thus, for older adults, situations that create negative stereotypical beliefs related to aging salience are likely to impact many health-related issues (Wurm et al., 2017).

One of the most investigated questions is how age-based stereotypes influence older people's cognitive performance (Armstrong et al., 2017; Horton et al., 2008; Lamont et al., 2015). Studies have shown that symbolic changes in the assessment

---

situation are sufficient to elicit a stereotype threat, which can hamper participants' performance (Barber, 2017). Generally, stereotype threat occurs when one "*must deal with the possibility of being judged or treated stereotypically, or when doing something that would confirm the stereotype*" (Steele & Aronson, 1998, p. 401). The term age-based stereotype threat (ABST) suggests that older adults' performance would be hampered in situations wherein they are expected to underperform because of their age. Early demonstrations were based on the characteristics of the task (Desrichard & Köpetz, 2005; Rahhal et al., 2001). For example, in Desrichard and Köpetz (2005), older participants performed worse on a task presented as "measuring memory abilities." Other studies directly manipulated stereotype salience. Hess et al. (2003) asked participants to read a fake newspaper article about the poorer cognitive performance of older adults versus one that contradicted common views about cognitive aging. Older participants in the negative condition performed worse on the recall task. Since then, several demonstrations of age stereotypes' effect on older adults' performance have been published (Barber, 2017) and meta-analyses have confirmed its robustness. Comparing positive and negative conditions, Horton et al. (2008) obtained an effect size of d= .38 across seven studies. Lamont et al. (2015) analyzed 53 effect sizes and concluded a mean effect size of d = .28. In their meta-analysis, Armstrong et al. (2017) examined the effects of stereotype threat on working memory and episodic memory performance. They obtained an average effect size of d = .373 (15 effect sizes) for the former and d = .253 (23 effect sizes) for the latter.

Several pathways and mechanisms of stereotypes' influence on cognitive performance have been discussed in the literature. The activation of stereotypes in a cognitive test situation can lower performance expectations (Desrichard & Köpetz, 2005). A lower performance expectation can lead a person to disengage more quickly from the task (i.e. to make less effort or to persist less; Beaudoin & Desrichard, 2017). Furthermore, the activation of a stereotype can generate stress and anxiety (Sindi et al., 2013) or a cognitive load that interferes with task performance (Mazerolle et al., 2012; Schmader et al., 2008; Schmader & Johns, 2003). Other authors propose that stereotype threat would lead older persons to adopt less effective strategies (Gimmig et al., 2006; Lemaire et al., 2018; Nicolas et al., 2020).

Can age difference activate a stereotype threat?

Meta-analyses have shown that subtle cues of age-related stereotypes are sufficient to affect older adults' cognitive performance (Armstrong et al., 2017; Horton et al., 2008; Lamont et al., 2015). A study by Kang et al. (Kang & Chasteen, 2009) showed that the presence of a young participant in the testing room could decrease older adults' performance. Older participants were tested in the presence of young adults versus older adults. The results showed that older adults performed worse on the free recall task when tested in the former condition. This effect also appeared in a cued recall task, but only among older adults who expressed a high level of perceived stereotype threat in the presence of younger adults. This is because younger participants evoke stereotype threats by making age categories more salient and reminding older adults of the expected differences in cognitive performance. Hence, the examiner's age could also influence older adults' performance due to similar mechanisms.

If a significantly younger examiner evaluates older people, they might be more aware of their age and, hence, feel more threatened by age stereotypes, whereas this would occur less when the examiner is older (i.e. closer to the participant's age). This prediction aligns with previous findings on stereotype threat induced by an examiner. For example, Marx and Goff (Marx & Goff, 2005) demonstrated that when black participants were tested by an experimenter of the same ethnic origin, their verbal performance did not deteriorate, whereas a white experimenter activated the stereotype of black Americans having inferior verbal capacities, resulting in lower verbal performance.

From a conceptual perspective, uncovering the age-of-examiner effect would provide new insights into the mechanisms driving older adults' cognitive performance. At a practical level, newly gained knowledge might help sensitize future examiners and medical and geriatric practitioners to assess older adults under fair and reliable conditions (Régner et al., 2016).

A useful source for tracking the age-of-examiner effect is data from large-scale surveys that include cognitive-screening tasks. They are often entrusted to survey agencies that offer a varied sample of interviewers, specifically, regarding age. When the examiner's age is reported, its relationship with the participant's performance can be tested, while controlling for potentially confounding variables. In the present study, we used data from two large-scale studies. Regarding previous literature on stereotype threat and the contextual malleability of older adults' test performance (Armstrong et al., 2017; Horton et al., 2008; Lamont et al., 2015), older adults' performance would increase as a function of the examiner's age, as older examiners would elicit less stereotype threat.

Material and method

Samples

We analyzed two independent samples to test our hypotheses. First, from the second wave of the Swiss National study, Vivre-Leben-Vivere (VLV; $N = 1250$; Oris et al., 2016). VLV is a multidisciplinary project aiming to study older adults' well-being (Laera et al., 2021). A survey was designed, and data collection was entrusted to the Link Institute, a leading Swiss polling company. It included measures of health, personality, cognition, social and human capital, lodging conditions, income, and wealth. Data were collected in 2017 from the participants' homes. We identified 960 participants (48.8% women, aged 70 to 102 years, $M = 80.62$, $SD = 6.62$) living at home with an MMSE score= 21, for whom we had examiners' age. Among them, 5.5% were single, 59.9% were in a partnership, 9.3% were separated, and 25.3% were widowed. The interviewers were employees of the survey agencies who were recruited and trained to conduct face-to-face interviews. We identified 29 different examiners (48% women) who had completed at least four interviews (M = 22.79, range 10–49, SD = 9.69). All interviewers had at least one year of experience (M = 6.48, SD = 5.60, range 1–27). Their age ranged between 19 and 79 years (M = 51.89, SD = 16.21).

The second sample comes from the Survey of Health, Ageing, and Retirement in Europe (SHARE; $N = 68231$; Börsch-Supan, 2020). SHARE is a European multidisciplinary

research project that aims to study health, social, economic, and environmental policies over the life course of 28 European countries and Israel. SHARE data collection is based on computer-assisted personal interviews (CAPI). Data were collected between September and December 2015. As each country uses its survey agency to collect data, it may lead to variability across countries. However, data are collected by interviewers employed by private agencies and trained in face-to-face interviews. We identified 32,768 participants (55.5% women, aged 51 to 105 years, $M=67,55$; $SD=10,01$) from 12 countries, and for whom we had the examiners' age. Among them, 69.4% were in a partnership, 12.3% were never married or divorced, and 18.2% were widowed. No initial cognitive screening for SHARE was performed. We identified 776 different examiners (69% women) who had conducted at least four interviews ($M=43.29$, aged 4 to 251 years, SD = 33.53) and had a mean experience of 10.49 years (aged 0 to 56 years, SD = 9.46). The education level of the interviewers was distributed as follows: lower-level secondary school, 5.6%; medium-level secondary school, 22.4%; lower-level secondary school, 28.7%; university degree, 43.3%). Their ages ranged from 18 to 78 years ($M=50.80$, SD = 13.03).

For both samples, the target sample comprises people aged 50 years or older who regularly reside in the country. Regarding sample selection, strict random sampling rules were imposed on survey agencies. The samples were stratified to represent the target population. Potential participants were contacted, and the volunteers were interviewed at home in their native language (French or German for VLV, and according to their country for SHARE). Concerning interviewers, the survey agencies did not provide detailed information on their training process. It can be speculated that they received general training in the interviews and data collection. The cognitive tasks were administered as part of the general questionnaire and were, therefore, not by professional neuropsychologists or psychologists. However, the instructions and procedures were precisely explained, and the tasks were computer-guided, thus ensuring an acceptable consistency level of consistency.

Cognitive measures

Cognitive measures were not the same across the samples. In the VLV, we used some of the subtests of the Cognitive Telephone Screening Instrument (COGTEL) to assess different cognitive functions among adults (Kliegel et al., 2007). The different cognitive tests used in SHARE were inspired by existing batteries (Dewey et al., 2005).

VLV—Prospective memory
These are four event-based tasks, a third of which was inspired by the version used in COGTEL, and the other three were added for this study. Throughout the interview, participants had to perform a predefined action immediately following a particular event: say, "red pencil" out loud when the interviewer mentions the red pencil, tap twice on the table when asked about physical activity, give their date of birth when asked about the activities performed during their life and the number of years of schooling, and remind the interviewer to turn on their cell phone at the end of the session. The score corresponded to the number of actions

correctly performed, without any reminders from the interviewer, and ranged from zero to four.

VLV—Working memory (backward digit recall)

The task comprised recalling sets of digits in reverse order of their presentation. Two trials per difficulty level were administered, presenting a series ranging from two digits from the easiest level to seven digits from the most difficult level. The interviewer read a series of words at a rate of one digit per second. When the participant failed both trials at the same level, the test ended. The score corresponds to the number of correctly reproduced sets of digits and ranges from zero to 12.

VLV—phonemic and semantic verbal fluency

This task combined phonemic and semantic fluency. Participants were asked to name, without repetition, 1) as many words as possible beginning with the letter A (excluding proper nouns) and 2) as many occupations as possible (the same occupations mentioned once in the feminine and masculine were counted as one) in one minute for each of these subtests. The score was the sum of the correct words produced for both.

VLV—Reasoning

Participants had to complete a logical series of five numbers (read at a rate of one digit every three seconds) with the appropriate number. The difficulty level increased with each trial. The test ended when the participant failed two consecutive trials. The score corresponds to the number of correctly completed logical series and ranged from zero to eight.

SHARE—Working memory (subtraction task)

Participants were asked to count backward from 100 by subtracting seven each time, five times in a row (i.e. the first subtraction was from 100 and then from the last number mentioned). If an error occurred in one trial, the answer given by the participant was considered to indicate if the next trial was correct. The score corresponds to the number of correctly performed subtractions and was between zero and five.

SHARE—Immediate recall

Participants were asked to memorize one of five lists of 10 words randomly assigned to them. The words were read aloud by an interviewer. Participants were then asked to recall as many words as possible immediately after the encoding phase. The score corresponded to the number of words recalled and ranged from zero to 10.

SHARE—Delayed recall

Participants were asked to recall the words on the list in a delayed manner after other cognitive tests were administered (which lasted about 10 min). The score corresponded to the number of words recalled and ranged from zero to 10.

SHARE—Semantic verbal fluency

Contrary to the verbal fluency task used in the VLC sample, this task had only a semantic component. Participants were asked to name as many animals as possible in one minute (excluding repetitions and proper names). The score represents the number of correct words that were produced.

Control variables

Experience in conducting interviews properly is a confounding variable. A better interviewer experience could lead participants to score higher, which would be a trivial explanation for the interviewer's age effect. To address this, we controlled for indicators of interviewer experience in the analyses: the examiner's experience in interviewing (in years) and the total number of interviews conducted during the study. We also controlled for the education level of the SHARE examiners, data that were not available in the VLV sample, and interviewers' and participants' sex.

Analysis

Listwise deletion was used for missing cases, yielding a slight variation in sample size depending on the task. The adjusted relationships between examiners' ages and participants' performance were estimated using multilevel regression. For the VLV data, a hierarchical structure was used wherein 960 participants were nested within 29 examiners, and the examiners were nested within two Swiss linguistic regions. We used a hierarchical structure for the SHARE data where 32768 participants were nested within 776 examiners, and the examiners were nested within 12 countries. We included random intercepts for the examiners and regions/countries. The examiners' sex, and age, and participants' sex were modeled as fixed effects for both samples. Fixed effects also included confounders that may be correlated with the examiners' age and reflect their interviewing ability: the examiner's interviewing experience (in years) and the total number of interviews conducted during the study. We also included the education level of SHARE examiners and other data that were not available in the VLV sample. We used standardized estimates of the slope (β) as estimates of the effect size. We expected a positive linear relationship between examiners' age and participants' performance, independent of other predictors. In Figure 1, we demonstrate the scale of the ordinate regarding the standard deviation quantity, which allows a better understanding of the effect size.

Results

Figure 1 depicts the relationships between interviewer's age and participants' performance levels predicted by multilevel regression, which includes interviewer's age, sex, and experience in interviewing, the total number of interviews conducted during the study, education level (SHARE only), and participant's sex. The figure shows a robust pattern of association between examiners' age and the performance predicted by the model. The single contribution of age (controlled for the other covariates) ranges
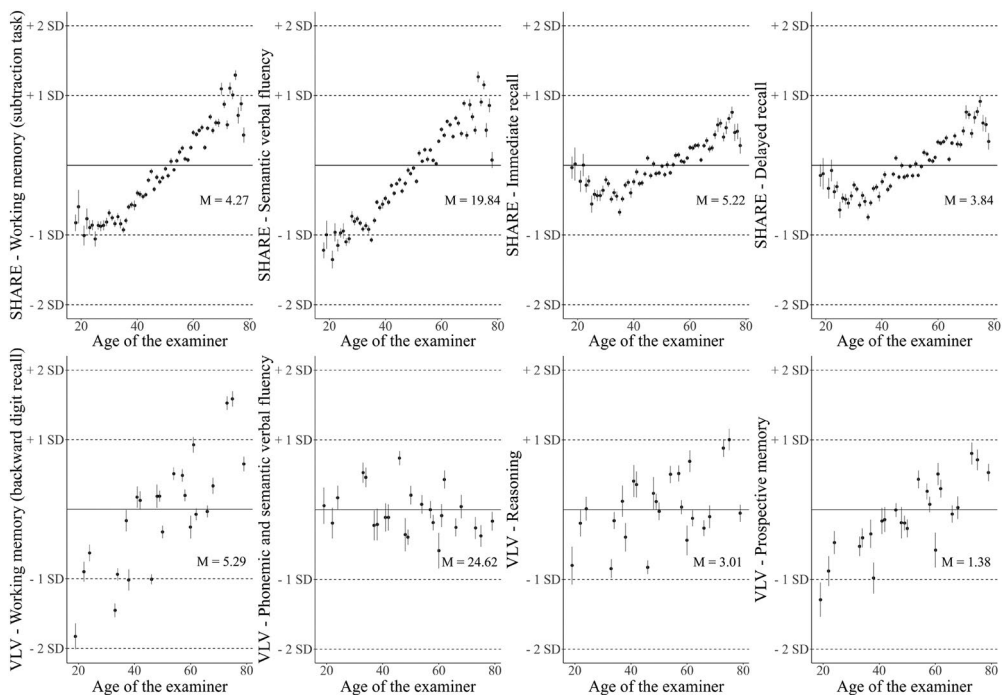
Figure 1. Estimated means of cognitive performance according to examiners' age. Bars indicate the standard error of the mean. N_{SHARE} = 32768, N_{VLV} = 960. M is the mean estimated performance across the groups.

from β = .0.05 to β = .274, indicating a small effect size according to Cohen's taxonomy (Cohen, 2013). The results reached significance for one of the four VLV tasks and all four SHARE tasks. After controlling for the other predictors, examiners' age was positively correlated with VLV participants' working memory score, b = .016, t(26,546) = 2.28, p=0.03, β = .151; SHARE participants' delayed recall score, b = .028, t(790,792) = 9.46, p=0,0001, β = .157; SHARE immediate recall score, b = .018, t(793,563) = 7.99, p=0.0001, β = .121; SHARE verbal fluency score, b = .16, t(782,006) = 13.25, p=0.0001, β = .274; and SHARE working memory (subtraction task) score, b = .013, t(733,924) = 9.48, p=0.0001, β = .157. The effect of examiners' age on VLV participants' performance did not reach significance for prospective memory score, b = .010, t(35,039) = 1.50, p=0,14, β = .133, reasoning, b = .01, t(29,888) = 1.014, p=0.31, β = .069. and verbal fluency, b=−.03, t(34,168) = −.85, p=0.40, β = −.057.

Discussion

Although multiple studies have established that variations in the assessment environment can affect older adults' cognitive performance, no study has tested the age-of-examiner effect. The age of an examiner is correlated with several characteristics, such as experience or skills, that can potentially influence participants' performance. Figure 1 shows that examiner's age, as a proxy for other variables, has a robust influence on participants' performance. The purpose of this study was to assess

whether examiner's age, independently from other covariates, could have a unique influence on participants' performance. We justified this hypothesis based on stereotype threat theory. Our results indicate that, after controlling for covariates, participants' performance on five tasks out of eight, was positively related to the examiners' age, showing that the older the examiner, the better the participant's performance.

This aligns with the results of Horton's meta-analysis (Horton et al., 2008), which shows a robust effect of age-based stereotype threat (ABST) threat on working and episodic memory performance. We observed an effect of experimenter age on delayed recall, which did not reach significance in Horton's meta-analysis (Horton et al., 2008). Lamont (Lamont et al., 2015) also showed that the effect of ABST is weaker when performance is measured at a distance from threat induction. This limitation does not apply to our study because the threat is related to the experimenter's presence and, hence, it is always present. This may explain why we observed an effect at the time of the delayed recall. Similarly, Horton (Horton et al., 2008) showed that the effect of ABST is significant on working memory when stereotype threat is manipulated blatantly, and on episodic memory when stereotype threat is manipulated subtly. We observed an effect of ABST on tasks involving working and episodic memory. This questions the status of the examiners' age from the perspective of an induced threat (blatant or subtle). Future studies could further explore this point by examining whether a threat related to the examiners' age is explicitly experienced by participants.

We observed a positive relationship with the prospective memory task performed by VLV participants, but it did not reach the significance level. This may reflect a lack of statistical power in the VLV sample because of the small number of different examiners. Moreover, participants might not identify this task as a memory assessment. Previous studies have shown that presenting a task that does not measure memory is sufficient to reduce the stereotype threat effect (Desrichard & Köpetz, 2005). We observed a null effect on the reasoning task and the VLV verbal fluency score was associated with significantly low or negative betas. It can be assumed that if there is a stereotype of cognitive decline with age, it is less pronounced in reasoning and verbal skills. Older adults could be less accepting of the popular idea of decline in these areas, although they have internalized the idea of memory decline. This may protect them from a stereotype threat effect on these skills. However, the verbal fluency task showed the examiners' age effect on the SHARE sample. Although the two tasks are different and, therefore, make a comparison between studies questionable, this lack of consistency may be a concern. In Switzerland, the stereotype associated with a decline in verbal fluency could be less prevalent than in other countries included in the SHARE sample, which may reduce the potential stereotype threat effect and induce a less robust effect on this skill.

A robust result is that the age-of-examiner effect is reliably present in memory performance (working memory, delayed recall, immediate recall, and, to some extent, prospective memory). Memory is strongly associated with the stereotype of aging. This is consistent with a stereotype threat effect, in that it is more likely to appear on cognitive tasks where the idea of decline is strongly present and internalized. The absence of results on the reasoning task could appear to be contradictory to this

conclusion, as it relies heavily on working memory. However, it is likely that the participants did not identify it as such.

These findings have important conceptual and practical implications for scientific research and clinical memory assessment. Scientific laboratories studies are typically conducted by younger examiners, which—as our data shows—negatively affects older adults' memory performance. Consequently, it is plausible that the age-related decline in memory performance observed by previous studies (Salthouse, 2009) may be—to a certain degree—overestimated because of the age-of-examiner effect. Similarly, a biased performance due to the age-of-examiner effect could cause some systematic errors. For example, if research aims to investigate factors that can enhance older adults' memory performance, the potential positive effects can (unwantedly) be hidden if the examiners' age hampers performance.

However, a better understanding of the age-of-examiner effect is also relevant in clinical settings. Although evaluating older adults' cognitive performance is a common practice in neuropsychology, gerontology, and medicine, realizing the age-of-examiner effect is important for providing a fair and reliable assessment. In these fields, older adults (for example, older people with suspected mild cognitive impairment) are sometimes examined by significantly younger individuals (by definition, professionals below retirement age), which may induce stereotype threat and hamper their performance. Biased assessment results can have significant consequences for providing proper diagnoses, suggesting appropriate treatment, and individuals' self-perception and well-being. It would be unrealistic to recommend that neuropsychological testing should be performed by older neuropsychologists. A simpler recommendation would be to be attentive to these aspects and try to not make age significantly salient in the testing situation (e.g. by avoiding asking whether the person has grandchildren or other age-marker questions). Another recommendation is to propose corrections in the tests according to the examiners' age. Future neuropsychological tests could incorporate this aspect into the score calculation.

The limitations of our study may have affected our conclusions. First, the age-of-the-examiner effect was small, which is reassuring regarding its clinical significance. Although some people might show sensitivity to examiners' age, which could strongly interfere with their performance and cause bias, group-level results suggest that it is unlikely that this effect is massive and widespread. Second, although participants' allocation to examiners is likely quasi-random, the samples were drawn from cross-sectional surveys, which do not offer the same guarantees as laboratory studies or randomized controlled trials. Third, the generalization of our results to neuropsychological assessment situations should be performed cautiously. The tasks used were not robust and were mostly useful for screening purposes. Furthermore, the examiners were not professional neuropsychologists. These aspects can significantly increase the age-of-examiner effect. Conversely, interviewing participants at home may help decrease the stereotype threat effect compared to a more stressful situation (Schlemmer & Desrichard, 2018; Strickland-Hughes & West, 2021). Therefore, studies conducted in a neuropsychological context are necessary to evaluate the effect of the examiner's age in this situation. Fourth, the agencies were responsible for allocating participants to the interviewers. It is questionable whether the allocation was random or guided by criteria that could bias the results (e.g. older people with memory problems

assigned to the most experienced interviewers). However, it is unlikely that this was the case as the agencies did not know the participants because they were randomly selected. Therefore, they did not have the opportunity (or probably desire) to assign them to interviewers based on any characteristic. Finally, we referred to stereotype threat to predict the effect of examiners' age. This interpretation assumes some form of causality (the situation elicits a stereotype threat that induces a decline in performance), which cannot be fully supported by correlational data. Furthermore, alternative and complementary explanations can be proposed. For example, older examiners could display behaviors (e.g. less anxiety and slower speech rates) that promote participants' performance. Examining the effect of examiners' age in experimental situations will lead to a better understanding of its mechanisms and some basic recommendations (e.g. speaking slowly) that could be included in the training of young neuropsychologists.

Beyond the limits that call for future investigations, this study confirms that subtle elements present in the evaluation of older adults can influence the scores obtained on cognitive assessments. Professional neuropsychologists have set up optimal conditions to obtain the most objective assessments. However, it is unclear whether all of them are aware of the potential impact that stereotypes can have on the performance of older adults. Future research should investigate the effects of subtle stereotype prompts in a clinical setting.

Acknowledgments

Authors' contributions

Conceptualization: OD; Methodology: OD, MK, OR; Investigation: OD, MO, MK; Visualization: NH; Funding acquisition: MK, MO; Project administration: OD, MK; Supervision: OD; Writing-original draft: OD; Writing-review & editing: OD, SZ, NH; Writing-revision: OD

Disclosure statement

Authors declare that they have no competing interests.

Funding

Data availability statement

Data and analysis scripts in the SPSS 26 format are available upon request.

ORCID

Matthias Kliegel ⓘ http://orcid.org/0000-0002-2001-2522

References

Armstrong, B., Gallant, S. N., Li, L., Patel, K., & Wong, B. I. (2017). Stereotype threat effects on older adults' episodic and working memory: A meta-analysis. *The Gerontologist*, *57*(suppl_2), S193–S205. https://doi.org/10.1093/geront/gnx056

Barber, S. J. (2017). An examination of age-based stereotype threat about cognitive decline: Implications for stereotype-threat research and theory development. *Perspectives on Psychological Science*, *12*(1), 62–90. https://doi.org/10.1177/1745691616656345

Barber, S. J., Hamel, K., Ketcham, C., Lui, K., & Taylor-Ketcham, N. (2020). The effects of stereotype threat on older adults' walking performance as a function of task difficulty and resource evaluations. *Psychology and Aging*, *35*(2), 250–266. https://doi.org/10.1037/pag0000440

Beaudoin, M., & Desrichard, O. (2017). Memory self-efficacy and memory performance in older adults: The mediating role of task persistence. *Swiss Journal of Psychology*, *76*(1), 23–33. https://doi.org/10.1024/1421-0185/a000188

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., & Zuber, S. (2013). Data resource profile: The survey of health, ageing and retirement in Europe (SHARE). *International Journal of Epidemiology*, *42*(4), 992–1001. https://doi.org/10.1093/ije/dyt088

Börsch-Supan, A. (2020). *Survey of health, ageing and retirement in Europe (SHARE) Wave 1 (7.1.0) [Data set]*. SHARE–ERIC. https://doi.org/10.6103/SHARE.W1.710

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge. https://doi.org/10.4324/9780203771587

Desrichard, O., & Köpetz, C. (2005). A threat in the elder: The impact of task-instructions, self-efficacy and performance expectations on memory performance in the elderly. *European Journal of Social Psychology*, *35*(4), 537–552. https://doi.org/10.1002/ejsp.249

Dewey, M., Prince, M., Börsch-Supan, A., Brugiavini, A., Jürges, H., Mackenbach, J., Siegrist, J., & Weber, G. (2005). *Health, ageing and retirement in Europe first results from the survey of health, ageing and retirement in Europe*. Mannheim Research Institute for the Economics of Aging.

Fresson, M., Dardenne, B., Geurten, M., & Meulemans, T. (2017). The effect of stereotype threat on older people's clinical cognitive outcomes: Investigating the moderating role of dementia worry. *The Clinical Neuropsychologist*, *31*(8), 1306–1328. https://doi.org/10.1080/13854046.2017.1307456

Gimmig, D., Huguet, P., Caverni, J.-P., & Cury, F. (2006). Choking under pressure and working memory capacity: When performance pressure reduces fluid intelligence. *Psychonomic Bulletin & Review*, *13*(6), 1005–1010. https://doi.org/10.3758/BF03213916

Hess, T. M., Auman, C., Colcombe, S. J., & Rahhal, T. A. (2003). The impact of stereotype threat on age differences in memory performance. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *58*(1), P3–P11. https://doi.org/10.1093/geronb/58.1.P3

Horton, S., Baker, J., Pearce, G. W., & Deakin, J. M. (2008). On the malleability of performance: Implications for seniors. *Journal of Applied Gerontology*, *27*(4), 446–465. https://doi.org/10.1177/0733464808315291

Kang, S. K., & Chasteen, A. L. (2009). The moderating role of age-group identification and perceived threat on stereotype threat among older adults. *The International Journal of Aging & Human Development*, *69*(3), 201–220. https://doi.org/10.2190/AG.69.3.c

Kliegel, M., Martin, M., & Jäger, T. (2007). Development and validation of the Cognitive Telephone Screening Instrument (COGTEL) for the assessment of cognitive function across adulthood. *The Journal of Psychology*, *141*(2), 147–170. https://doi.org/10.3200/JRLP.141.2.147-172

Laera, G., Joly-Burra, E., Zuber, S., Ballhausen, N., Künzi, M., Ihle, A., da Silva Coelho, C., Haas, M., Mikneviciute, G., Tinello, D., Kliegel, M., & Hering, A. (2021). Do executive functions explain older adults' health-related quality of life beyond event-based prospective memory? *Aging, Neuropsychology, and Cognition*, 1–15. https://doi.org/10.1080/13825585.2021.1989368

Lamont, R. A., Swift, H. J., & Abrams, D. (2015). A review and meta-analysis of age-based stereotype threat: Negative stereotypes, not facts, do the damage. *Psychology and Aging*, *30*(1), 180–193. https://doi.org/10.1037/a0038586

Lemaire, P., Brun, F., & Régner, I. (2018). Negative aging stereotypes disrupt both the selection and execution of strategies in older adults. *Gerontology*, *64*(4), 373–381. https://doi.org/10.1159/000486756

Levy, B. R. (2003). Mind matters: Cognitive and physical effects of aging self-stereotypes. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *58*(4), P203–P211. https://doi.org/10.1093/geronb/58.4.P203

Levy, B. R. (2009). Stereotype embodiment: A psychosocial approach to aging. *Current Directions in Psychological Science*, *18*(6), 332–336. https://doi.org/10.1111/j.1467-8721.2009.01662.x

Marx, D. M., & Goff, P. A. (2005). Clearing the air: The effect of experimenter race on target's test performance and subjective experience. *The British Journal of Social Psychology*, *44*(Pt 4), 645–657. https://doi.org/10.1348/014466604X17948

Mazerolle, M., Régner, I., Morisset, P., Rigalleau, F., & Huguet, P. (2012). Stereotype threat strengthens automatic recall and undermines controlled processes in older adults. *Psychological Science*, *23*(7), 723–727. https://doi.org/10.1177/0956797612437607

Nicolas, P., Lemaire, P., & Régner, I. (2020). When and how stereotype threat influences older adults' arithmetic performance: Insight from a strategy approach. *Journal of Experimental Psychology. General*, *149*(2), 343–367. https://doi.org/10.1037/xge0000647

Oris, M., Guichard, E., Nicolet, M., Gabriel, R., Tholomier, A., Monnot, C., Fagot, D., & Joye, D. (2016). Representation of vulnerability and the elderly. A total survey error perspective on the VLV survey. In M. Oris, C. Roberts, D. Joye, & M. Ernst Stähli (Éds.), *Surveying human vulnerabilities across the life course* (Vol. 3, pp. 27–64). Springer International Publishing. https://doi.org/10.1007/978-3-319-24157-9_2

Rahhal, T. A., Hasher, L., & Colcombe, S. J. (2001). Instructional manipulations and age differences in memory: Now you see them, now you don't. *Psychology and Aging*, *16*(4), 697–706. https://doi.org/10.1037/0882-7974.16.4.697

Régner, I., Mazerolle, M., Alescio-Lautier, B., Clarys, D., Michel, B., Paccalin, M., Piolino, P., Rigalleau, F., Sambuchi, N., & Huguet, P. (2016). Aging stereotypes must be taken into account for the diagnosis of prodromal and early Alzheimer disease. *Alzheimer Disease and Associated Disorders*, *30*(1), 77–79. https://doi.org/10.1097/WAD.0000000000000129

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, *30*(4), 507–514. https://doi.org/10.1016/j.neurobiolaging.2008.09.023

Schlemmer, M., & Desrichard, O. (2018). Is medical environment detrimental to memory? A test of a white coat effect on older people's memory performance. *Clinical Gerontologist*, *41*(1), 77–81. https://doi.org/10.1080/07317115.2017.1307891

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, *85*(3), 440–452. https://doi.org/10.1037/0022-3514.85.3.440

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*(2), 336–356. https://doi.org/10.1037/0033-295X.115.2.336

Sindi, S., Fiocco, A. J., Juster, R.-P., Pruessner, J., & Lupien, S. J. (2013). When we test, do we stress? Impact of the testing environment on cortisol secretion and memory performance in older adults. *Psychoneuroendocrinology*, *38*(8), 1388–1396. https://doi.org/10.1016/j.psyneuen.2012.12.004

Steele, C. M., & Aronson, J. (1998). Stereotype threat and the test performance of academically successful African Americans. In C. Jencks & M. Phillips (Eds.), *The Black–White test score gap* (pp. 401–427). Brookings Institution Press.

Strickland-Hughes, C. M., & West, R. L. (2021). The impact of naturalistic age stereotype activation. *Frontiers in Psychology*, *12*, 685448. https://doi.org/10.3389/fpsyg.2021.685448

VanLandingham, H., Ellison, R. L., Laique, A., Cladek, A., Khan, H., Gonzalez, C., & Dunn, M. R. (2022). A scoping review of stereotype threat for BIPOC: Cognitive effects and intervention strategies for the field of neuropsychology. *The Clinical Neuropsychologist*, *36*(2), 503–522. https://doi.org/10.1080/13854046.2021.1947388

Wurm, S., Diehl, M., Kornadt, A. E., Westerhof, G. J., & Wahl, H.-W. (2017). How do views on aging affect health outcomes in adulthood and late life? Explanations for an established connection. *Developmental Review: DR*, *46*, 27–43. https://doi.org/10.1016/j.dr.2017.08.002