PRIMENJENA PSIHOLOGIJA

No2, 2025

Contents

Primenjena psihologija
Vol. 18, No. 2, pp. 175–204, 2025

Research Article

Psychometric characteristics of the Serbian version of the Ethical Leadership at Work Questionnaire (ELW-RS)

Nikola Goljović[1]

[1] *Departmant of Psychology, Faculty of Philosophy, University of Niš, Serbia*

ABSTRACT

The aim of this study was to adapt the Ethical Leadership at Work Scale (ELW) into Serbian as a self-report measure for subordinates to evaluate their managers, to examine the construct validity of this new version, and to test the level of invariance across gender and organizational sector within the Serbian version (ELW-SR). The sample consisted of 306 participants (72.5% females) with at least six months of work experience, aged between 20 and 63 years, mostly highly educated, and primarily employed on a permanent basis in the private sector (74.80%). The results showed that ELW-SR adequately reflects all dimensions of the original scale: people orientation, fairness, power sharing, concern for sustainability, ethical guidance, role clarification, and integrity. Confirmatory factor analysis showed that a seven-dimensional solution with correlated residuals has a good fit (χ^2 = 1491.09, df = 640; CFI = .916, TLI = .907, RMSEA = .07, and SRMR = .100), thereby affirming the validity of the Serbian version of the scale. However, the study points to the need for caution in generalizing results, particularly concerning variables such as gender and type of organization, where the assumptions of invariance were not unequivocally confirmed. Additionally, the scale's criterion validity was tested to examine how well the test results predict relevant outcomes related to ethical leadership, such as job satisfaction, psychological safety, and self-efficacy. These findings imply that ELW-SR can be an effective tool for assessing ethical leadership in the Serbian business context.

Keywords: ethical leadership, ELW, CFA, invariance, seven-dimensional model

UDK: 159.938.3
DOI: 10.19090/pp.v18i2.2550
Received: 05.05.2024.
Revised: 25.07.2024.
Accepted: 29.10.2024.

✉ Corresponding author's email: nikolagoljovic@gmail.com

Introduction

The importance of studying ethical leadership within business environments is increasingly evident in modern society, which is striving more and more for transparency, accountability, and adherence to ethical norms in all aspects of business (Banks et al., 2019). Although research on ethical leadership has primarily been conducted in Western countries, particularly the USA, where business ethics measures are well established by legislation, there is a distinct need to expand the research focus to various cultural and regional contexts (Saha et al., 2020). This is particularly significant in Central and Eastern European (CEE) countries, where leadership studies are not as advanced as in the West, making such research crucial for understanding the nuances and challenges faced by this region (Boţa-Avram et al., 2021).

Given the significant interest in ethical leadership in recent scholarship, the current conceptualization of ethical leadership reveals substantial limitations that impede both theoretical and practical advancements (Banks et al., 2021; Fischer et al., 2020; Saha et al., 2020). The prevalent models conflate leader behaviors with followers' subjective evaluations, incorporating an array of traits, values, and cognitions that may not accurately represent ethical leadership behaviors (ELB). Moreover, the causal relationships between ethical leadership and its outcomes remain obscured due to methodological shortcomings in existing studies (Banks et al., 2021). This conceptual confusion necessitates a refined measurement approach to capture the multifaceted nature of ethical leadership more accurately.

The cultural context in which leadership is studied also can significantly affect the portrayal and perception of ethical behaviors within the leadership context. Despite the existence of initial intercultural comparisons (Eisenbeiß, 2012; Eisenbeiß & Brodbeck, 2013), it has been essential to expand research to develop a comprehensive view of ethical leadership, including aspects that are universally accepted and aspects specific to certain cultures. Applying this approach to the Serbian context and validating the Serbian version of the Ethical Leadership at Work questionnaire are important steps towards better understanding both global and local dynamics of ethical leadership. Ethical Leadership at Work questionnaire not

only contributes to the academic community with new, culturally relevant insights but also enables organizations within the CEE region to better assess and develop ethical leadership practices, which is crucial for building sustainable and ethically responsible business practices.

With this goal, the Ethical Leadership at Work questionnaire (ELW) by Kalshoven and colleagues (Kalshoven et al., 2011) became a focus of interest for researchers. By adapting the multidimensional Ethical Leadership at Work questionnaire into Serbian, our aim was to provide a comprehensive and valid multidimensional scale in the Serbian language, which can help us address various issues of importance in this area.

Understanding Ethical Leadership

As per Brown et al. (2005), ethical leadership can be described as "the demonstration of normatively appropriate conduct through personal actions and interpersonal relationships, and the advancement of such conduct to the followers through two-way communication, reinforcement, and decision making" (p. 120). This approach identifies two key components: the leader as a moral person and the leader as a moral agent, emphasizing the necessity for the leader to be authentic in their moral principles before implementing these principles in leading others (Treviño et al., 2000). The dimension of a of a leader as a moral person is reflected in traits such as integrity, fairness, and authenticity, which are evident both in their professional and private lives. These characteristics are not unique to ethical leadership but also overlap with other leadership styles, such as authentic, servant, transformational, and spiritual leadership (Toor & Ofori, 2009). The component of a leader as a moral agent particularly pertains to the leader's efforts to promote ethical values among followers, encouraging them to behave ethically and make ethical decisions. This role involves developing a culture that values transparency, accountability, and mutual respect, directly impacting the perception of the leader as an ethical guide (Brown & Treviño, 2006; Treviño et al., 2003).

Previous empirical studies have shown a significant association between ethical leadership and the ethical behavior of employees. Specifically, "ethical behavior of employees" refers to actions such as adherence to company policies, reporting misconduct, treating colleagues

with respect, maintaining honesty in communications, and demonstrating fairness and integrity in decision-making processes (Brown et al., 2005; Resick et al., 2006; Ofori & Toor, 2021). Moreover, ethical leadership is associated with enhanced interactions between leaders and followers, contributing to improved exchange quality and leadership efficiency (Den Hartog & Belschak, 2012; Hassan et al., 2013).

Research has demonstrated the crucial role ethical leaders play in shaping and enhancing an organization's ethical climate (Schminke et al., 2005; Mayer et al., 2009; Kang et al., 2011). Such leaders establish the ethical norms of an organization through the implementation of processes and policies that influence employees' perceptions of their work environment's ethicality (Demirtas et al., 2015; Mayer et al., 2009). In environments like Serbia, business practices exhibit unique characteristics, especially in leadership approaches and ethical standards. Serbian business culture has historically included hierarchical structures, with authoritative and directive leadership styles often observed, especially in the context of transitional economic and political challenges (Mojić, 2003). While Serbian business culture has a legacy of self-governing socialism—traditionally emphasizing community and employee well-being—leadership practices have shifted notably in recent decades (Hollinshead & Maclean, 2007). Ethically motivated leadership and people-oriented practices in Serbia face unique challenges rooted in the complex post-socialist landscape. As shown in a study analyzing the fragmented narratives in a Serbian enterprise recently acquired by a multinational company, the volatile institutional and politically charged context complicates the applicability of linear, Western models of organizational change (Hollinshead & Maclean, 2007). This context amplifies the need for an in-depth exploration of ethical leadership within Serbian organizations, as it highlights the potential for both ethical and unethical practices influenced by socio-political dynamics and transitional realities. Schminke et al. (2005) found that the interplay between a leader's moral standards and organizational factors—including caring orientation, regulatory orientation, and autonomous decision-making—correlates strongly with the organization's ethical climate. Consequently, further studies underscore the critical impact of ethical leadership on fostering an ethical organizational environment (Lu & Lin, 2014; Mayer et al., 2009).

Ethical leaders demonstrate respect for their followers, supporting and caring for them, consistent with findings by Treviño et al. (2003). They are distinguished by a high people orientation, as indicated by Eisenbeiß and Brodbeck (2013). By including followers' ideas and concerns in their decisions, allowing participation in setting performance goals, and giving followers a voice, ethical leaders engage in power sharing, which contributes to the development of followers and strengthens their self-confidence (Brown et al., 2005; Kalshoven et al., 2011). Research also indicates that ethical leadership practices can impact employees' levels of self-efficacy, leading to heightened engagement and dedication to organizational objectives (Tongsoongnern & Lee, 2022). Moreover, the presence of ethical leaders in the workplace can cultivate a sense of optimism among employees, encouraging their commitment to the organization's success (Hoogh & Hartog, 2008).

Fairness is also a key characteristic of ethical leaders, as suggested by the moral dimension of personality, reflected through transparent, objective, and balanced decisions and interactions (Treviño et al., 2000). Integrity is manifested through alignment of behavior with stated principles, keeping promises, and consistency in actions (Brown & Treviño, 2006; Kalshoven et al., 2011). Ethical leaders exhibit a broad ethical awareness that transcends organizational boundaries, particularly evident in their deep concern for sustainability, including care for the environment and promoting environmentally friendly work processes (Kalshoven et al., 2011). Ethical guidance involves explaining the values and guidelines of ethics to followers, emphasizing the importance of ethical standards, and making ethics an explicit part of the leadership agenda (Brown & Treviño, 2006). Through ethical guidance, leaders reward those who act according to ethical standards and penalize those who violate them, fostering ethical awareness among followers.

While the dimensions of people orientation, fairness, and integrity constitute the moral person within ethical leadership (Brown et al., 2005), power sharing, role clarification, ethical guidance, and concern for sustainability form the facet of moral management (Khuntia & Suar, 2004). Such a comprehensive definition has enabled a deeper understanding and

measurement of ethical leadership, providing a foundation for further research and practical application in organizations.

Measuring Ethical Leadership

The first instruments began to develop in the early 21st century. Brown and colleagues (Brown et al., 2005) developed the Ethical Leadership Scale (ELS), which has been widely used in research. However, ELS faces criticism due to the breadth and ambiguity of its items, as well as a lack of precision in defining behaviors related to ethical leadership (Tanner et al., 2010). These weaknesses have highlighted the need for the development of new instruments that better capture the multidimensionality of ethical leadership, including a leader's personal moral traits and managerial practices.

In developing the Ethical Leadership at Work Scale (ELW; Kalshoven et al., 2011), the authors aimedto overcome the limitations of previous instruments (e.g., the ELS), focusing on more precise item formulation and an expanded set of dimensions encompassing ethical leadership. The ELW is based on extensive analysis of existing research, interviews with managers and employees, and original items developed by the researchers, using an empirical-descriptive approach. This approach emphasizes the concrete behaviors of leaders and their interactions with employees, allowing for the assessment of ethical leadership without prior knowledge of ethics. The evaluation is based on the frequency of displaying certain behaviors rather than on assessing the ethicality of leaders' actions, highlighting the interaction between leaders and followers as key to perceiving ethical leadership.

The ELW enhances its predecessor, the Ethical Leadership Scale (ELS), by focusing on moral management aspects such as power sharing, role clarification, and sustainability. It has been adapted into multiple languages, reflecting its global relevance and the broad interest it has generated. The ELW not only expands the definition of fairness to include daily work interactions but also integrates environmental considerations, aligning with corporate social responsibility and business ethics. Additionally, it assesses leaders' integrity but omits trust, offering a comprehensive tool for evaluating ethical leadership (Kalshoven et al., 2011).

The ELW not only represents a theoretical breakthrough but also excels in practical applications, providing a comprehensive framework for analyzing the various dimensions of ethical leadership (Kalshoven et al., 2011; Silva & Duarte, 2022). The ELW scale encompasses several dimensions of ethical leadership, including fairness, integrity, ethical guidance, people orientation, power sharing, role clarification, and concern for sustainability. These dimensions are interrelated and collectively contribute to a comprehensive understanding of ethical leadership. For instance, fairness and integrity are foundational elements that support trust and respect within an organization, while ethical guidance and people orientation foster a supportive and morally sound work environment (Kalshoven et al., 2011).

Empirical studies using the ELW scale have measured various work attitudes, including job satisfaction, organizational commitment, and trust in leadership (Jang & Oh, 2017; Kim & Park, 2015). For instance, ethical leadership significantly enhances job satisfaction by promoting a positive work environment and fair treatment (Kalshoven et al., 2011; Steinmann et al., 2016). Organizational commitment, reflecting employees' emotional attachment and loyalty to their organization, is also positively influenced by ethical leadership, which fosters a sense of belonging and ethical culture (Kim & Park, 2015; Metwally et al., 2019). Additionally, the ELW concept incorporates sustainability as a crucial construct (Kalshoven et al., 2011), particularly predicting green behavior (Ahmad et al., 2021). The scale has also been predictive of other important outcomes, such as employee engagement and reduced counterproductive work behaviors (Huang et al., 2021). Ethical leaders who exhibit behaviors such as fairness, integrity, ethical guidance, and concern for sustainability create an environment where employees feel psychologically safe to voice their opinions and concerns (Ahmad & Umrani, 2019). Ethical leadership has been found to enhance psychological safety by fostering trust, respect, and open communication within the organization (Walumbwa & Schaubroeck, 2009).

The ELW scale is positively related to other leadership styles, such as transformational leadership transactional or severant leadership, and negatively related to autocratic and passive leadership (Steinmann et al., 2016; Zhu et al., 2015). Both ethical and transformational leadership

emphasize the moral and inspirational aspects of leadership, though ethical leadership places a stronger emphasis on ethical conduct and moral principles (Den Hartog, 2015). Ethical leadership provides incremental value by specifically addressing ethical issues and promoting a culture of integrity, which may not be as explicitly covered in transformational leadership (Chun et al., 2009; Den Hartog, 2015).

Study Aims

The objective of this study was to adapt the Ethical Leadership at Work Scale (ELW; Kalshoven et al., 2011) into Serbian (ELW-SR), examine the construct validity of this new version, and assess its alignment with the original ELW scale; our aim was to offer a new and more comprehensive self-report measure of employees' perceptions of ethical leadership of their superiors. This study builds on the constructs used by Kalshoven and colleagues (2011) to validate ELW. While adapting the ELW to Serbian, particular emphasis was placed on tailoring the scale to reflect the cultural and linguistic nuances of the region. This adaptation involved not only linguistic accuracy in translation but also ensuring that the concepts of ethical leadership were relevant and comprehensible within the Serbian workplace environment. Additionally, the study investigated the factor structure of the ELW-SR, verifying whether the original seven-dimensional structure remained stable in the Serbian context. Assessing the factor structure was crucial for confirming the structural validity of the scale and its dimensions across different cultural settings. Gender and organizational type (public vs. private) were tested for measurement invariance. Previous studies indicated that male and female leaders may exhibit different leadership styles shaped by societal expectations and traditional values (Mitrić-Aćimović et al., 2012; Stojanović-Aleksić et al., 2016; Stošić Panić & Simić, 2024). Additionally, public sector organizations face unique ethical challenges (due to their bureaucratic structures and reform pressures), and thus, they may differ from the private sector (Janovac et al., 2023). By examining scale invariance across gender and organizational type, this study aims to ensure that the scale accurately captures employees' perceptions of ethical leadership in diverse contexts, enhancing the validity and applicability of the measure across various demographic and professional groups in Serbia. We examined the criterion validity of ELW-SR by analyzing its

correlations with psychological safety, self-efficacy, and job satisfaction, which are considered outcomes of ethical leadership. These constructs were chosen based on their established relevance in leadership research.

Method

Sample

The initial sample consisted of 392 employees. Given that the scale measures the ethical behavior of managers as rated by employees, we excluded individuals with less than six months of work experience, those without a direct supervisor, and those not part of a team of at least three people. This refinement left a sample of 312 individuals. After removing multivariate outliers based on Mahalanobis distance criteria (Tabachnick & Fidell, 2007), 306 participants remained. We used a convenience sample, comprising individuals with at least six months of work experience (Min = 0.7, Max = 37, *M* = 7.96, *SD* = 9.09), of which 72.5% were female. Participants' age ranged form 20 to 63 years (*M* = 31.85, *SD* = 8.62). The highest percentage of participants had higher education, with completed bachelor's (32.4%) or master's degrees (36.9%), while 18.6% had finished vocational high school or gymnasium, and the smallest percentages had completed associate degrees (6.9%) or doctoral studies (5.2%). Nearly 90% of participants came from urban areas, 5.6% from towns, and 5.2% from rural areas. In terms of employment, 56.9% were on permanent contracts, 36.3% on temporary contracts, and 6.9% employed on other bases. A majority of 74.8% worked in private organizations, while 25.2% were in the public sector.

Measures

Ethical Leadership at Work Questionnaire (ELW; Kalshoven et al., 2011)

The ELW was designed to explore the prerequisites and outcomes of ethical leadership by asking subordinates to rate their supervisors' ethical leadership behaviors. Participants were instructed to read each item carefully and decide the extent to which they agreed or disagreed with it using a five-point Likert scale (1 - *strongly disagree*, 5 - *strongly agree*). The ELW features 38 items spread across seven dimensions: People Orientation ("Is

genuinely concerned about my personal development"), Fairness ("Holds me accountable for problems over which I have no control" *reverse item), Power Sharing ("Allows subordinates to influence critical decisions"), Sustainability Concern ("Shows concern for sustainability issues"), Ethical Guidance ("Explains what is expected from employees in terms of behaving with integrity"), Role Clarification ("Indicates what the performance expectations of each group member are"), and Integrity ("Keeps his/her promises"). The initial version's reliability in original study ranged from .84 to .94. For this study, the scale was translated into Serbian using a back-translation method. Two bilingual translators translated and re-translated the 38 items, discussing and reconciling differences to agree on a functionally equivalent Serbian version.

Perceived Organizational Support Scale (POSS; Armstrong-Stassen & Ursel, 2009)

Originally developed by Eisenberger et al. (1986), this scale measures perceived organizational support and was adapted by Armstrong-Stassen and Ursel. It consists of 10 items, with a five-point Likert scale as a response format (1 - *strongly disagree*; 5 - *strongly agree*). An example item is: "The organization values my contribution to its well-being.".This unidimensional questionnaire was adapted to Serbian using a back-translation method for this research. Original studies have shown it to have strong metric properties, with a Cronbach's alpha coefficient of internal consistency of .95.

Psychological Safety Questionnaire (PSQ; Edmondson, 1999, adaptation Goljović, 2023)

The PSQ is a unidimensional questionnaire consisting of 7 items, with responses also on a five-point Likert scale (1 - *strongly disagree*; 5 - *strongly agree*). Respondents need to answer each question based on their personal experience in the current work environment/team. An example item is: "I feel safe to take a risk in this organization.". It was adapted to Serbian using a back-translation method for this study. The scale has demonstrated good metric characteristics in original research, with a Cronbach's alpha coefficient of internal consistency of .80.

Job Satisfaction Measure (JSS; Dolbier et al., 2005)

This single-item questionnaire asks respondents to rate their job satisfaction on a five-point Likert scale (1 - not at all satisfied; 5 - extremely satisfied), answering the question: "Overall, how satisfied are you with your job?" This single-item measure is designed to measure the general affective dimension of job satisfaction. Initial research demonstrated satisfactory reliability and validity of this measure, and the justification for its use has been supported in numerous studies (Ock, 2020).

Procedure and Data Analysis

Following permission from one of the authors of the original scale for its use and translation, a forward translation was conducted by two independent translators (Hedrih, 2019), followed by data collection. Data analysis involved confirmatory factor analysis (CFA) using the maximum likelihood method with IBM SPSS and the AMOS (version 21, extension), which was used to assess the structural validity of the instrument. Specifically, we aimed to assess whether the factor structure of the Serbian adaptation of the scale corresponds to the original factor structure. Both a unidimensional (single-factor) model and a seven-factor model, including a variant with correlated residuals, were tested. Model fit was evaluated using various indices, including $\chi 2$, $\chi 2/df$, Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). A model was considered to have an acceptable fit if CFI and TLI values were .90 or higher, and if the values of RMSEA and SRMR were .08 or lower (Kline, 2011). Criterion validity was tested to examine how well the test results predict relevant outcomes related to ethical leadership. Invariance was tested against two criteria: gender and the type of organization in which participants work (i.e., employed in the public or private sector).

Results

The results of the CFA (Table 1) indicated that the unidimensional model had a relatively high χ^2/df ratio, with fit indices (i.e., CFI, TLI, RMSEA, and SRMR) suggesting inadequate model fit.

The seven-factor model showed improvement over the unidimensional model. However, although the fit of this model was better (compared to the unidimensional model), the fit indices still fell short of the acceptability thresholds. Significant improvement was achieved with the introduction of correlated residuals into the seven-factor model. This model displayed a χ^2/df of 2.330, CFI of .916, and TLI of .907, indicating a good model fit. The RMSEA value reduced, also suggesting a good fit. However, the SRMR value remained significantly above acceptable limits.

Table 1

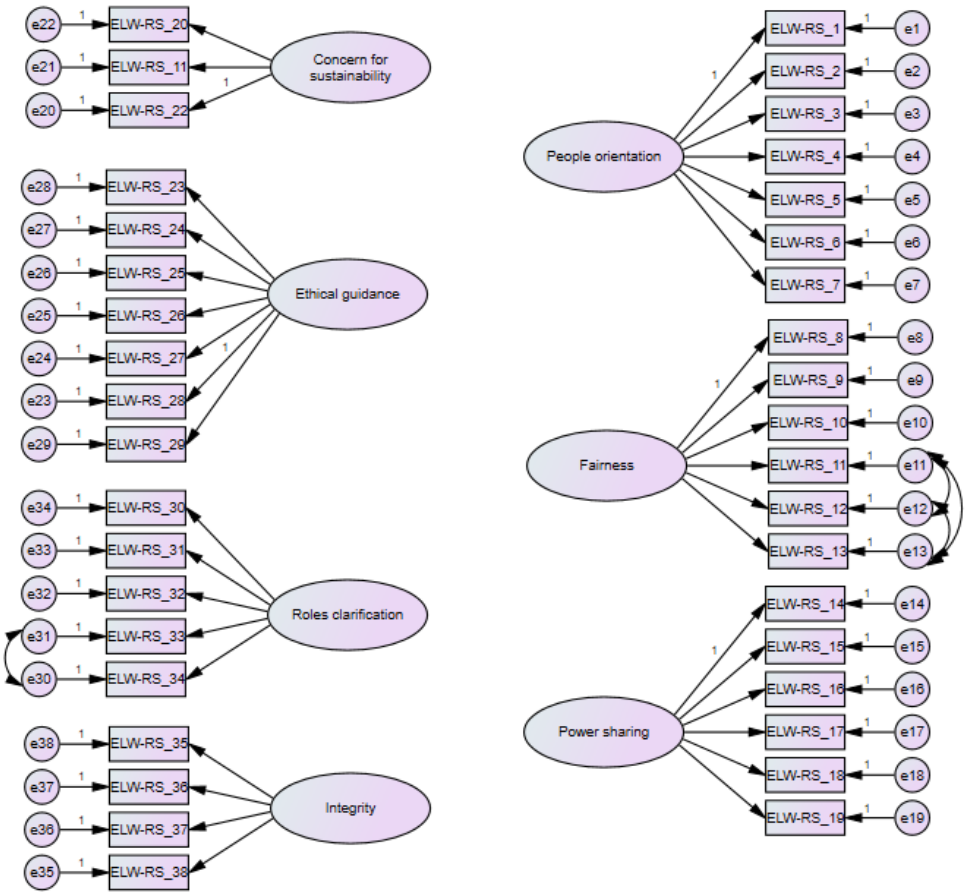
Results of Confirmatory Factor Analysis

	χ^2	*df*	χ^2/*df*	*CFI*	*TLI*	*RMSEA*	*SRMR*
1-factor model	2076.53	659	3.153	.859	.850	.084	.237
7-factor model	1769.46	644	2.748	.888	.878	.080	.091
7-factor model with correlated residuals	1491.09	640	2.330	.916	.907	.066	.100

An ANOVA test was conducted to compare the fit of different models. The differences between the models were assessed using ANOVA tests of the chi-square values. The results indicated that the difference between the unidimensional model and the 7-factor model was statistically significant ($\Delta\chi^2 = 307.07$, $\Delta df = 15$, $p < .001$), suggesting that the 7-factor model provides a significantly better fit than the unidimensional model. Additionally, the difference between the 7-factor model and the 7-factor model including correlated dimenions and several residuals was also statistically significant ($\Delta\chi^2 = 278.37$, $\Delta df = 4$, $p < .001$), indicating that including correlations significantly improves the model fit. Furthermore, the comparison between the unidimensional model and the 7-factor model with correlated dimenions revealed a significant difference ($\Delta\chi^2 = 585.44$, $\Delta df = 19$, $p < .001$), providing further support for the best fit of the 7-factor model with correlated dimenions (for the graphical representation of this model, see Figure 1).

Figure 1

A Seven-factor Solution with Correlated Dimensions and Residuals



Note. Correlations between latent dimensions were included in the analysis but not shown in the model for clarity.

The next segment of the analysis focuses on testing the model's invariance with respect to two key demographic criteria: gender and the type of organization (public or private). Invariance is crucial for determining whether the factor structures of the model are consistent across the studied groups. The analysis was conducted at four levels: configural, metric, scalar, and strict invariance, each imposing increasingly stringent conditions of equality among groups. The results are summarized in tables that illustrate

how the model performs in terms of different fit indices (χ^2/df, CFI, TLI, RMSEA) and changes in these indices (ΔCFI, ΔTLI, ΔRMSEA) across the various levels of invariance.

The analysis of invariance by gender (Table 2) indicated that the model meets the basic adaptability requirements across all levels of invariance. Although the CFI (.879) and TLI (.867) values were relatively high, they did not reach the commonly recommended thresholds (.90) for optimal model fit. These values suggested acceptable, but not ideal, fitting, implying that the model adequately represented the data structure in relation to gender. Changes in fit indices (ΔCFI, ΔTLI, ΔRMSEA) between different levels of invariance were minimal, suggesting that the model maintained consistency in measurement across both genders.

Table 2

The Analysis of Invariance across Gender

	χ^2/*df*	*CFI*	*TLI*	*RMSEA*	*Δ CFI*	*Δ TLI*	*Δ RMSEA*
Configural	2.00	.879	.867	.067	-	-	-
Metric	1.98	.889	.867	.067	.00	.00	.00
Scalar	1.95	.889	.867	.067	.00	.00	.00
Strict	1.95	.890	.868	.066	.01	.00	.00

The analysis of invariance by organizational type (public vs. private) (Table 3) showed that models at different levels of invariance had decent, but not ideal, adaptability indices. The configural model had a CFI of .883 and TLI of .872, indicating a fit below the usual threshold of .90. The CFI and TLI values were similar across models. Such findings suggested that the factor structures remained relatively consistent regardless of the considered levels of invariance. RMSEA values were consistently low (.054-.056); however, the CFI and TLI values were below the recommended thresholds.

Table 3

The Analysis of Invariance across Type of Organization (Private or Public)

	χ^2/df	*CFI*	*TLI*	*RMSEA*	*Δ CFI*	*Δ TLI*	*Δ RMSEA*
Configural	1.94	.883	.872	.056	-	-	-
Metric	1.92	.883	.872	.056	.00	.00	.00
Scalar	1.92	.881	.871	.056	.00	.00	.00
Strict	1.90	.881	.870	.054	.00	.00	.00

In the next step, we assessed the interconnectedness between various dimensions of ethical leadership, including orientation to people, fairness, power sharing, sustainability, ethical guidance, role clarification, and integrity (Table 4). All correlations were statistically significant and moderate in magnitude, implying that different aspects of ethical leadership were closely linked. Such findings indicated the conceptual coherence of the ethical leadership construct as a whole, supporting the internal validity of the measured dimensions.

The reliability of each dimension (Table 4), assessed by Cronbach's alpha coefficients, had values ranging from .814 to .958, suggesting an exceptionally high degree of internal consistency for each dimension.

Table 4

Descriptive Statistics, Pearson Correlation Coefficients, and Reliability Measures

	1	2	3	4	5	6	7
People orientation	(.94)						
Fairness	.41**	(.89)					
Power sharing	.60**	.33**	(.81)				
Concern for sustainability	.41**	.18**	.37**	(.89)			
Ethical guidance	.60**	.32**	.40**	.42**	(.92)		
Roles clarification	.66**	.25**	.46**	.28**	.67**	(.93)	
Integrity	.73**	.49**	.47**	.36**	.62**	.62**	(.96)
M	3.68	4.00	3.56	2.94	3.65	3.83	3.93

SD	1.06	1.09	.87	1.23	1.11	1.03	1.10
Sk	-.57	-1.02	-.43	.04	-.63	-.80	-.89
Ku	-.67	.11	-.31	-.95	-.38	-.20	-.07

Note. Sk - skewness; Ku - kurtosis; values in parentheses indicate Cronbach alpha coefficients.

$^{*}p < .05$. $^{**}p < .01$.

The results also revealed significant positive correlations between job satisfaction and all dimensions of ethical leadership, with correlations ranging from moderate to high (r = .31 - .52). These correlations support the criterion validity of the instrument. The relationships between self-efficacy and ethical leadership were weaker; still, self-efficacy showed significant correlations with all dimensions of ethical leadership except for fairness. Psychological safety showed moderate to high significant correlations with all dimensions of ethical leadership, suggesting that a greater perception of ethical leadership contributes to a greater sense of safety among employees.

Table 5

Pearson Correlation Coefficients between Ethical Leadership and Related Constructs

	People orientation	Fairness	Power sharing	Concern for sustainability	Ethical guidance	Roles clarif.	Integrity
Job satisfaction	.51**	.31**	.45**	.31**	.42**	.43**	.52**
Self-efficacy	.24**	.05	.21**	.14*	.20**	.15*	.21**
Psy. safety	.52**	.38**	.40**	.28**	.31**	.35**	.47**

$^{*}p < .05$. $^{**}p < .01$.

Discussion

The aim of this study was to translate the Ethical Leadership at Work Scale (ELW) into Serbian, to examine the construct validity of the new version, and to demonstrate the level of invariance between the Serbian version of the ELW (ELW-SR). This study builds on the constructs proposed by Kalshoven et al. (2011), who developed the Ethical Leadership at Work scale as an extension and elaboration of the earlier ethical leadership framework introduced by Brown et al. (2005). We verified whether the original seven-dimensional structure remains stable in translation.

Consistent with the ELW, the Serbian version of this instrument (ELW-SR) comprises the dimensions of people orientation, fairness, power sharing, concern for sustainability, ethical guidance, role clarification, and integrity. Fit indices confirm that the seven-dimensional solution provides the most optimal fit. Comparative model analyses suggest that the seven-dimensional solution is the most suitable model among those tested. Similar to the initial version, values indicating that a one-factor solution would be adequate were not obtained. Even though the initial validation yielded significantly better results, the one-factor solution was not retained as the final one nor was it recommended to use a composite score as a unitary measure of ethical leadership. The structure of ELW-SR effectively reflects the key aspects of ethical leadership, as identified in the original study. Since residual correlations were introduced for items that belong to the same measurement subject and conceptually describe very similar phenomena, this solution was retained as meaningful and adopted as final.

The invariance of the scale was examined in relation to two criteria: gender and type of organization, which enabled the determination of whether the scale measures ethical leadership equally well among different demographic and professional groups in Serbia. The CEE region has distinct historical, social, and economic characteristics that influence organizational behavior and leadership styles (Cartwright, 2020). For instance, the legacy of hierarchical and authoritative leadership styles from the pre-transition period may impact how ethical leadership is perceived and practiced (Csath, 2022). Therefore, it is crucial to validate and adapt the ELW-SR to ensure it accurately reflects the ethical leadership constructs within this specific cultural milieu. By doing so, we can ensure that the instrument is sensitive to

cultural nuances and provides valid and reliable measurements of ethical leadership in the CEE context.

The decision to test for invariance across different genders and types of organizations stems from theoretical and empirical considerations. Gender differences in leadership have been widely documented in the literature, with research suggesting that men and women may exhibit and perceive leadership behaviors differently (Ho et al., 2015; Kacmar et al., 2011). Eagly and Johnson (1990) posited that women tend to adopt a more transformational leadership style, characterized by empathy and ethical considerations, whereas men may lean towards transactional leadership. Given these differences, it is essential to examine whether the ELW-SR is equally valid for both genders to ensure it does not inadvertently favor one over the other. Additionally, the variation in organizational types—such as public vs. private sectors—can also influence the practice and perception of ethical leadership. Public sector organizations often emphasize transparency and accountability, while private sector entities might prioritize efficiency and profitability (Andersen, 2010). These differing organizational cultures can shape the way ethical leadership is enacted and perceived. By testing the invariance of the ELW-SR across different types of organizations, we can ascertain its robustness and applicability in varied organizational settings, ensuring its broader utility and relevance.

However, the results of the present study do not unequivocally indicate invariance. A positive aspect of the obtained results is that there are no differences in versions according to both criteria. The differences in fit indices that were obtained are negligible. Nevertheless, the fit indices achieved in this analysis are not within acceptable limits but are slightly below. Although such results could be accepted according to some less stringent criteria, the conclusion is that this solution is not acceptable and does not positively support the invariance of the measure relative to the examined criteria.

Criterion validity was further explored through a comparative analysis of ELW-SR with existing instruments measuring related outcomes of leadership, ensuring that ELW-SR adequately reflects the role of ethical leadership on important workplace outcomes. Supporting the scale's criterion validity, our results revealed significant positive correlations

between job satisfaction and all dimensions of ethical leadership. The association with the dimension of self-efficacy, as well as psychological safety, shows medium to high significant correlations with all dimensions of ethical leadership, suggesting that a higher perception of ethical leadership correlates with a greater sense of personal efficacy and a higher sense of security among employees within the same team. The results support the idea that all seven dimensions should be considered when studying ethical leadership. This provides additional value in predicting outcomes and contributes to a more detailed understanding of how ethical leadership develops or functions, as well as of particularly effective behaviors. The results are consistent with previous research examining how ethical leadership can contribute to job satisfaction (Jang & Oh, 2017; Steinmann et al., 2016), psychological safety (Ahmad & Umrani, 2019), and ultimately, employees' self-efficacy (Hoogh & Hartog, 2008; Tongsoongnern & Lee, 2022).

Within our study, significant limitations were identified that deserve special attention in order to deepen the understanding of the results obtained and guide future research initiatives. One of the key limitations was related to the sample of respondents, which was not sufficiently balanced in terms of variables relevant for measuring invariance, which is the most significant flaw observed in the process of assessing psychometric characteristics. This imbalance may have contributed to the less favorable results; thus, further research is needed to address this aspect in detail and draw reliable conclusions. Furthermore, there is a pronounced need for more thorough research into the nomological network of leadership to provide a clearer understanding of this complex construct. Additionally, given the closeness of the construct of ethical leadership to related concepts such as servant and transformational leadership, it is advisable to conduct an analysis of the instrument's discriminative validity. Such an analysis is crucial for making a reliable decision about the psychometric characteristics of the scale, ensuring its validity and reliability in academic and practical applications.

From a theoretical perspective, this research contributes to the literature on ethical leadership by expanding the empirical basis of the ELW scale and testing its applicability in a different cultural context. The findings affirm that the concept of ethical leadership is relevant and applicable

beyond the Anglo-Saxon context, offering insights into the universality and cultural specifics of ethical leadership. Moreover, the study highlights the importance of further examining and validating the multidimensionality of ethical leadership.

The practical implications of this research are substantial, particularly in the context of fostering ethical leadership within Central and Eastern European (CEE) organizations. The ELW-SR instrument provides a robust tool for organizations seeking to elevate ethical standards and encourage ethical behavior in the workplace. This instrument enables leaders and HR professionals to identify specific areas for development and training, thereby facilitating targeted interventions to enhance ethical leadership. Furthermore, employing the ELW-SR can help cultivate a work environment characterized by transparency, integrity, and fairness. By regularly assessing ethical leadership behaviors, organizations can build and sustain a strong ethical culture, which in turn can lead to increased employee satisfaction and loyalty. Ultimately, the utilization of the ELW-SR has the potential to improve overall organizational efficiency by ensuring that ethical considerations are integral to leadership practices.

Conclusion

The objective of this study was to translate the Ethical Leadership at Work Scale into Serbian and to examine the factor structure, measurement invariance, and construct validity of the Serbian version. The results demonstrated that the seven-dimensional structure of the scale effectively reflects the key aspects of ethical leadership, despite challenges related to invariance across gender and organizational type. Overall, the instrument shows promising psychometric properties, and the author recommends its further use, emphasizing the scale's importance in the development and research of ethical leadership, particularly in the demographic area where the validation was conducted. This instrument not only provides insights into the specific dimensions of ethical leadership but also encourages organizations to actively engage in promoting ethical values and behaviors, which is crucial for building sustainable and responsible business practices.

Conflict of interest

We have no conflicts of interest to disclose.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Ahmad, I., & Umrani, W. A. (2019). The impact of ethical leadership style on job satisfaction: Mediating role of perception of Green HRM and psychological safety. *Leadership & Organization Development Journal*, *40*(5), 534-547. https://doi.org/10.1108/lodj-12-2018-0461

Ahmad, S., Islam, T., Sadiq, M., & Kaleem, A. (2021). Promoting green behavior through ethical leadership: a model of green human resource management and environmental knowledge. *Leadership & Organization Development Journal*, *42*(4), 531-547. https://doi.org/10.1108/LODJ-01-2020-0024

Andersen, J. A. (2010). Public versus private managers: How public and private managers differ in leadership behavior. *Public Administration Review*, *70*(1), 131-141. https://doi.org/10.1111/j.1540-6210.2009.02117.x

Armstrong-Stassen, M., & Ursel, N. D. (2009). Perceived organizational support, career satisfaction, and the retention of older workers. *Journal of Occupational and Organizational Psychology, 82*(1), 201–220. https://doi.org/10.1348/096317908X288838

Banks, G. C., Fischer, T., Gooty, J., & Stock, G. (2021). Ethical leadership: Mapping the terrain for concept cleanup and a future research agenda. *The Leadership Quarterly*, *32*(2), 101471. https://doi.org/10.1016/j.leaqua.2020.101471

Banks, G. C., Woznyj, H. M., Wesslen, R. S., Frear, K. A., Berka, G., Heggestad, E. D., & Gordon, H. L. (2019). Strategic recruitment across borders: An investigation of multinational enterprises. *Journal of Management, 45*(2), 476-509. https://doi.org/10.1177/0149206318764295

Boţa-Avram, C., Groşanu, A., & Răchişan, P. R. (2021). Investigating country-level determinant factors on ethical behavior of firms: evidence from CEE countries. *Journal of East-West Business*, *27*(2), 184-205. https://doi.org/10.1080/10669868.2020.1869638

Brown, M. E., and Treviño, L. K. (2006). Ethical leadership: a review and future directions. *Leadersh. Q. 17*, 595–616. https://doi.org/10.1016/j.leaqua.2006.10.004

Brown, M. E., Treviño, L. K., and Harrison, D. (2005). Ethical leadership: a social learning perspective for construct development and testing. *Organizational Behavior and Human Decision Processes, 97*, 117–134. https://doi.org/10.1016/j.obhdp.2005.03.002

Cartwright, C. T. (2020). Understanding global leadership in eastern and central Europe: the impacts of culture and intercultural competence. In *Understanding National Culture and Ethics in Organizations* (pp. 63-74). Emerald Publishing Limited. https://doi.org/10.1108/978-1-83867-022-120201007

Chun, J. U., Yammarino, F. J., Dionne, S. D., Sosik, J. J., & Moon, H. K. (2009). Leadership across hierarchical levels: Multiple levels of management and multiple levels of analysis. *The Leadership Quarterly, 20*(5), 689–707. https://doi.org/10.1016/j.leaqua.2009.06.003

Csath, M. (2022). Leadership development: Examples from Central and Eastern Europe. In *Developing Leaders for Real: Proven Approaches that Deliver Impact* (pp. 43-53). Emerald Publishing Limited. https://doi.org/10.1108/978-1-80071-364-220221009

Demirtas, O. (2015). Ethical leadership influence at organizations: Evidence from the field. *Journal of Business Ethics, 126*(2), 273-284. https://doi.org/10.1007/s10551-013-1950-5

Den Hartog, D. N. (2015). Ethical leadership. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*(1), 409-434. https://doi.org/10.1146/annurev-orgpsych-032414-111237

Den Hartog, D. N., & Belschak, F. D. (2012). Work engagement and Machiavellianism in the ethical leadership process. *Journal of Bussiness Ethics 107*, 35–47. https://doi.org/10.1007/s10551-012-1296-4

Dolbier, C. L., Webster, J. A., McCalister, K. T., Mallon, M. W., & Steinhardt, M. A. (2005). Reliability and validity of a single-item measure of job satisfaction. *American Journal of Health Promotion*, *19*(3), 194-198. https://doi.org/10.4278/0890-1171-19.3.194

Eagly, A. H., & Johnson, B. T. (1990). Gender and leadership style: A meta-analysis. *Psychological Bulletin*, *108*(2), 233. https://doi.org/10.1037/0033-2909.108.2.233

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*(2), 350-383. https://doi.org/10.2307/2666999

Eisenbeiß, S. A. (2012). Re-thinking ethical leadership: an interdisciplinary integrative approach. *Leadersh. Q. 23*, 791–808. https://doi.org/10.1016/j.leaqua.2012.03.001

Eisenbeiß, S. A., and Brodbeck, F. (2013). Ethical and unethical leadership: A cross-cultural and cross-sectional analysis. *Journal of Bussiness Ethics. 122*, 343–359. https://doi.org/10.1007/s10551-013-1740-0

Fischer, T., Hambrick, D. C., Sajons, G. B., & Van Quaquebeke, N. (2020). Beyond the ritualized use of questionnaires: Toward a science of actual behaviors and psychological states. *The Leadership Quarterly*, *31*(4), 101449. https://doi.org/10.1016/S1048-20)30076-X

Goljović, N. (2023). The role of emotional development of employees and people oriented leadership in perceiving psychological safety in organizations. *Godišnjak za Psihologiju, 20*, 75-88. https://doi.org/10.46630/gpsi.20.2023.05

Hassan, S., Mahsud, R., Yukl, G., & Prussia, G. E. (2013). Ethical and empowering leadership and leader effectiveness. *Journal of Managerial Psychology, 28*(2), 133–146. https://doi.org/10.1108/02683941311300252

Hedrih, V. (2019). *Adapting Psychological Tests and Measurement Instruments for Cross-Cultural Research: An Introduction (1st ed.)*. Routledge. https://doi.org/10.4324/9780429264788

Ho, S. S., Li, A. Y., Tam, K., & Zhang, F. (2015). CEO gender, ethical leadership, and accounting conservatism. *Journal of business ethics*, *127*, 351-370. https://doi.org/10.1007/s10551-013-2044-0

Hollinshead, G., & Maclean, M. (2007). Transition and organizational dissonance in Serbia. *Human Relations, 60*(10), 1551-1574. https://doi.org/10.1177/0018726707083477

Hoogh, A. H. B. D. and Hartog, D. N. D. (2008). Ethical and despotic leadership, relationships with leader's social responsibility, top management team effectiveness and subordinates' optimism: a multi-method study. *The Leadership Quarterly, 19*(3), 297-311. https://doi.org/10.1016/j.leaqua.2008.03.002

Huang, S. Y., Li, M. W., & Chang, T. W. (2021). Transformational leadership, ethical leadership, and participative leadership in predicting counterproductive

work behaviors: evidence from financial technology firms. *Frontiers in Psychology*, *12*, 658727. https://doi.org/10.3389/fpsyg.2021.658727

Jang, Y., & Oh, Y. (2019). Impact of ethical factors on job satisfaction among Korean nurses. *Nursing Ethics*, *26*(4), 1186-1198. https://doi.org/10.1177/0969733017742959

Janovac, T., Djokovic, G., Pusara, A., Misic, V., Milankovic, K., Pavicevic, A., Vukovic, A., & Jovanovic, S. V. (2023). Assessment and ranking of the behavioural leadership model in the process of implementing reforms in the public sector of the Republic of Serbia using the PIPRECIA method. *Sustainability, 15*(13), 10315. https://doi.org/10.3390/su151310315

Kacmar, K. M., Bachrach, D. G., Harris, K. J., & Zivnuska, S. (2011). Fostering good citizenship through ethical leadership: Exploring the moderating role of gender and organizational politics. *Journal of Applied Psychology*, *96*(3), 633-642. https://doi.org/10.1037/a0021872

Kalshoven, K., Den Hartog, D. N., and De Hoogh, A. H. B. (2011). Ethical leadership at work questionnaire: Development and validation of a multidimensional measure. *Leadersh. Q. 22*, 51–69. https://doi.org/10.1016/j.leaqua.2010.12.007

Khuntia, R., & Suar, D. (2004). A scale to assess ethical leadership of Indian private and public sector managers. *Journal of business ethics*, *49*, 13-26. https://doi.org/10.1023/B:BUSI.0000013853.80287.da

Kim, J. E., & Park, E. J. (2015). A validation study of the modified Korean version of ethical leadership at work questionnaire (K-ELW). *Journal of Korean Academy of Nursing*, *45*(2), 240-250. https://doi.org/10.4040/jkan.2015.45.2.240

Kline, R. B. (2013). Assessing statistical aspects of test fairness with structural equation modelling. *Educational Research and Evaluation, 19*(2-3), 204–222. https://doi.org/10.1080/13803611.2013.767624

Lu, C. S., & Lin, C. C. (2014). The effects of ethical leadership and ethical climate on employee ethical behavior in the international port context. *Journal of Business Ethics*, *124*, 209-223. https://doi.org/10.1007/s10551-013-1868-y

Metwally, D., Ruiz-Palomino, P., Metwally, M., & Gartzia, L. (2019). How ethical leadership shapes employees' readiness to change: The mediating role of an organizational culture of effectiveness. *Frontiers in Psychology*, *10*, 2493. https://doi.org/10.3389/fpsyg.2019.02493

Mitrić-Aćimović, D., Vujić, D., & Dostanić, J. (2012). Relations between assumptions of human nature and work and leadership style preferences. *Primenjena Psihologija*, *5*(4), 375–392. https://doi.org/10.19090/pp.2012.4.375-392

Mojić, D. (2003). The influence of National Culture on Organizational subcultures and Leadership styles in Serbian enterprises: an empirical analysis. *Sociologija*, *45*(4), 317-346.

Ock, J. (2020). How satisfied are you with your job? Estimating the reliability of scores on a single-item job satisfaction measure. *International Journal of Selection and Assessment*, *28*(3), 297-309. https://doi.org/10.1111/ijsa.12285

Ofori, G., & Toor, S.-U.-R. (2021). *Leadership in the Construction Industry: Developing Authentic Leaders in a Dynamic World* (1st ed.). Routledge. https://doi.org/10.1201/9781003038757

Resick, C. J., Hanges, P. J., Dickson, M. W., & Mitchelson, J. K. (2006). A cross-cultural examination of the endorsement of ethical leadership. *Journal of business ethics*, *63*, 345-359. https://doi.org/10.1007/s10551-005-3242-1

Saha, R., Shashi, Cerchione, R., Singh, R., & Dahiya, R. (2020). Effect of ethical leadership and corporate social responsibility on firm performance: A systematic review. *Corporate Social Responsibility and Environmental Management*, *27*(2), 409-429. https://doi.org/10.1002/csr.1824

Schminke, M., Ambrose, M. L., & Neubaum, D. O. (2005). The effect of leader moral development on ethical climate and employee attitudes. *Organizational behavior and human decision processes*, *97*(2), 135-151. https://doi.org/10.1016/j.obhdp.2005.03.006

Silva, V. H., & Duarte, A. P. (2022). Portuguese version of Brown, Treviño and Harrison's Ethical Leadership Scale: Study of its psychometric properties. *Cogent Business & Management*, *9*(1), 2153437. https://doi.org/10.1080/23311975.2022.2153437

Steinmann, B., Nübold, A., & Maier, G. W. (2016). Validation of a German Version of the Ethical Leadership at Work Questionnaire by Kalshoven et al. (2011). *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.00446

Stojanović-Aleksić, V., Stamenković, M., & Milanović, M. (2016). The analysis of leadership styles in Serbian organizations: Gender influence. *TEME –*

Časopis za Društvene Nauke, 40(4), 1383-1397. https://doi.org/10.22190/TEME1604383S

Stošić Panić, D., & Simić, I. (2024). Preferred leadership behavior in the Serbian setting: A cross-regional perspective. *Cross-Cultural Research, 58*(4), 358-369. https://doi.org/10.1177/10693971241262065

Stouten, J., van Dijke, M., Mayer, D. M., De Cremer, D., and Euwema, M. C. (2013). Can a leader be seen as too ethical? The curvilinear effects of ethical leadership. *Leadersh. Q. 24*, 680–695. https://doi.org/10.1016/j.leaqua.2013.05.002

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics (5th ed.)*. Allyn and Bacon.

Tanner, C., Brügger, A., van Schie, S., and Lebherz, C. (2010). Actions speak louder than words: Benefits of ethical behaviors of leaders. *Journal of Psychology, 218*, 225–233. https://doi.org/10.1027/0044-3409/a000032

Tongsoongnern, P. and Lee, W. S. (2022). Influence of green transformational leadership on the workplace pro-environment behavior. *Economic Analysis, 55*(2), 91-106. https://doi.org/10.28934/ea.22.55.2.pp91-106

Toor, S., & Ofori, G. (2009). Ethical leadership: Examining the relationships with full range leadership model, employee outcomes, and organizational culture. *Journal of Business Ethics, 90*, 533–547. https://doi.org/10.1007/s10551-009-0059-3

Treviño, L. K., Brown, M., and Hartman, L. P. (2003). A qualitative investigation of perceived executive ethical leadership: perceptions from inside and outside the executive suite. *Human Relations, 56*, 5–37. https://doi.org/10.1177/0018726703056001448

Treviño, L. K., Hartman, L. P., and Brown, M. (2000). Moral person and moral manager: How executives develop a reputation for ethical leadership. *California Management Review, 42*(4), 128−142. https://doi.org/10.2307/41166057

Walumbwa, F. O. and Schaubroeck, J. (2009). Leader personality traits and employee voice behavior: mediating roles of ethical leadership and work group psychological safety. *Journal of Applied Psychology, 94*(5), 1275-1286. https://doi.org/10.1037/a0015848

Zhu, W., Zheng, X., Riggio, R. E., & Zhang, X. (2015). A critical review of theories and measures of ethics-related leadership. *New Directions for Student Leadership*, *2015*(146), 81-96. https://doi.org/10.1002/yd.20137

Supplementary Materials

Table S1

Descriptive Statistics: Individual Items of the ELW-SR

	Item translated to Serbian language	*M*	*SD*	*Sk*	*Ku*	*Loading*
	Orijentacija na ljude					
ELW-RS_1	Zainteresovan je za to kako se zaista osećam i kako mi ide na poslu.	3.82	1.13	-.77	-.12	.86
ELW-RS_2	Ostavlja dovoljno vremena za uspostavljanje ličnog kontakta.	3.77	1.20	-.68	-.58	.80
ELW-RS_3	Obraća pažnju na moje potrebe.	3.70	1.21	-.65	-.57	.86
ELW-RS_4	Odvaja vreme za razgovor sa mnom o tome kako se osećam na poslu.	3.46	1.34	-.37	-1.12	.82
ELW-RS_5	Iskreno je zainteresovan za moj lični razvoj.	3.47	1.30	-.41	-.93	.81
ELW-RS_6	Saoseća sa mnom kad imam problema.	3.73	1.23	-.66	-.60	.82
ELW-RS_7	Brine o svojim podređenima.	3.84	1.21	-.75	-.46	.85
	Pravičnost					
ELW-RS_8	Smatra me odgovornim/om za probleme nad kojima nemam kontrolu.	3.93	1.18	-.92	-.14	.88
ELW-RS_9	Smatra me odgovornim/om za posao nad kojim nemam kontrolu.	3.95	1.20	-.98	-.07	.94
ELW-RS_10	Smatra me odgovornim/om za stvari koje nisu moja krivica.	4.13	1.14	-1.20	.42	.86
ELW-RS_11	Gradi sopstveni uspeh na rezultatima koji su postigli drugi.	3.73	1.41	-.70	-.90	.53
ELW-RS_12	Fokusiran je uglavnom na postizanje sopstvenih ciljeva.	3.30	1.46	-.29	-1.29	.47
ELW-RS_13	Manipuliše podređenima.	3.96	1.35	-1.03	-.28	.56

Podela moći						
ELW-RS_14	Omogućava podređenima da utiču na donošenje ključnih odluka.	3.25	1.19	-.40	-.67	.64
ELW-RS_15	Ne dozvoljava drugima da učestvuju u donošenju odluka.	3.95	1.19	-1.04	.21	.58
ELW-RS_16	Traži savete od podređenih koji se tiču organizacione strategije.	3.53	1.20	-.57	-.45	.74
ELW-RS_17	Preispituje odluke na osnovu preporuka koje dobije od svojih podređenih.	3.74	1.08	-.68	-.09	.70
ELW-RS_18	Dodeljuje podređenima zaduženja koja su izazovna.	3.80	1.13	-.92	.27	.54
ELW-RS_19	Dozvoljava mi da igram ključnu ulogu prilikom definisanja ciljeva i željenih rezultata rada.	3.48	1.23	-.51	-.64	.70
Briga za održivost						
ELW-RS_20	Želi da radimo na ekološki prihvatljiv način.	3.11	1.31	-.15	-.89	.81
ELW-RS_21	Pokazuje zabrinutost za pitanja održivosti životne sredine.	2.95	1.38	.03	-1.17	.94
ELW-RS_22	Stimuliše recikliranje predmeta i materijala u našem odeljenju.	2.74	1.39	.21	-1.17	.82
Etično vođenje						
ELW-RS_23	Jasno objašnjava koja su pravila ponašanja i kako da svi "igraju pošteno".	3.60	1.25	-.60	-.62	.50
ELW-RS_24	Objašnjava šta se od zaposlenih očekuje u pogledu etičnog ponašanja.	3.80	1.23	-.81	-.31	.84
ELW-RS_25	Pojašnjava dileme koje su u vezi sa etičnim ponašanjem na radnom mestu.	3.67	1.23	-.64	-.56	.89
ELW-RS_26	Brine se o tome da svi zaposleni prate etičke propise.	3.64	1.26	-.63	-.62	.92

ELW-RS_27	Pojašnjava moguće posledice neetičnog ponašanja mene i mojih kolega.	3.55	1.21	-.52	-.61	.89
ELW-RS_28	Podstiče raspravu među zaposlenima o pitanjima moralnog postupanja.	2.66	1.29	.26	-.93	.83
ELW-RS_29	Pohvaljuje zaposlene koji se ponašaju u skladu sa smernicama o radnoj etici i moralnom ponašanju.	3.42	1.33	-.43	-.92	.66
	Razjašnjenje uloga					
ELW-RS_30	Jasno ukazuje na to koji posao treba uraditi, za svakog člana tima.	3.63	1.17	-.58	-.43	.73
ELW-RS_31	Objašnjava šta se očekuje od svakog člana grupe.	3.73	1.22	-.65	-.68	.75
ELW-RS_32	Objašnjava šta se očekuje od mene i mojih kolega.	3.91	1.14	-.89	-.12	.94
ELW-RS_33	Pojašnjava prioritete.	4.01	1.15	-1.08	.36	.94
ELW-RS_34	Pojašnjava ko je za šta odgovoran.	3.88	1.17	-.93	.07	.81
	Integritet					
ELW-RS_35	Održava svoja obećanja.	3.92	1.12	-.79	-.13	.95
ELW-RS_36	Može mu se verovati da radi ono što kaže.	3.97	1.19	-1.00	.09	.97
ELW-RS_37	Moguće je pouzdati se u njega da će ispuniti ono što je obećao.	3.95	1.19	-.99	.07	.90
ELW-RS_38	Uvek drži svoju reč.	3.89	1.17	-.90	.04	.87

Primenjena psihologija
Vol. 18, No. 2, pp. 205–225, 2025

Research Article

The relation between different executive functions and depressive symptoms in young adults in Croatia

Tamara Bolić [1], Silvija Ručević [2], Sandra Brezetić [2]✉

[1] *Institute of Social Sciences Ivo Pilar, Croatia*
[2] *Department of Psychology, Faculty of Humanities and Social Sciences, J. J. Strossmayer University of Osijek, Croatia*

ABSTRACT

The study aims to explore the specific relation of different executive functions with depressive symptoms in a non-clinical sample of young adults in Croatia. The online study included 290 young adults (aged 18-35). Participants completed a PHQ-9 questionnaire that assessed depressive symptoms, one measure of executive functions using self-report, and two behavioural tasks measuring working memory and inhibition. Partial correlation analysis indicated that better working memory (measured using both self-report and a behavioural task) is related to less depressive symptoms in young adults. Contrary to expectations, the relation between inhibition and depressive symptoms was not confirmed. Additionally, the regression analysis suggests that working memory task is an important predictor of depressive symptoms, even after controlling for participant's education level, financial status, and level of resilience. Research shows that certain executive functions are differentially related to mental health in young adults. Additional research is needed to explore the mechanism underlying these differences and aid the creation of more appropriate treatment plans.

Keywords: executive functions, working memory, inhibition, depressive symptoms, young adults

UDK: 159.97.072-057.87(497.5)
DOI: 10.19090/pp.v18i2.2568
Received: 29.08.2024.
Revised: 18.10.2024.
Accepted: 29.10.2024.

✉ Corresponding author's email: svuckovic@ffos.hr

Introduction

People are often exposed to stressors that can lead either to adaptation or to various adverse health, behavioural, or psychological consequences. For example, it is estimated that, during their lifetime, approximately 70% of people will experience at least one traumatic life event such as exposure to death or the threat of death, serious physical injury (actual or threatened), or sexual violence (experienced directly, witnessed first-hand or indirectly, or experienced in professional activity; Frewen et al., 2019). A particularly challenging period that may require good adaptation skills is young adulthood. The period of transition from adolescence to adulthood often involves people changing their environment and exploring new experiences and new social roles. Various tasks that can arise in young adulthood make this period particularly stressful, especially for individuals with lower levels of psychological resilience. Among others, a particularly common psychiatric disorder among young adults is depression. In the United States, between 2015 and 2020, levels of depression in the population under 35 years of age increased, and 17.2% of young people aged 18 to 25reported experiencing at least one major depressive episode (Goodwin et al., 2022). Younger adults were more likely to report depressive symptoms than any other age group. Additionally, Arias-de la Torre et al. (2023) noted that the point prevalence of clinically relevant depressive symptoms in Europe between 2013 and 2020 had high variability across countries, with the highest within-country prevalence increases being observed in Slovenia, Denmark, Lithuania, and Croatia.

In addition to dissatisfaction, extreme feelings of sadness and pessimism, symptoms of depression involve the occurrence of physical, social, and cognitive changes (i.e., changes in sleeping habits and appetite, lack of motivation and energy, difficulty concentrating and ability to make decisions, and withdrawal from social activities). In individuals with major depressive disorder, all of these symptoms cause clinically significant distress or impairment in work, social, or other areas of functioning (American Psychiatric Association, 2013). Severe forms of depression may involve suicidal thoughts that can lead to suicide attempts and, consequently, deaths by suicide. Finally, many individuals who have subsyndromal depression may later develop difficulties that will require

intervention (Goodwin et al., 2022); therefore, it is crucial to recognize the factors that are associated with symptoms of depression in young adults.

Successful adaptation models emphasize a combination of physiological, neurobehavioral, environmental, and psychological factors. These psychological factors include executive functions, which play an important role in understanding complex and abstract concepts, solving new problems, and planning and governing different situations. A large body of research shows that as part of the developmental process of adaptation, undifferentiated neutral systems in childhood become increasingly specialized with age. For example, a well-known "unity and diversity" theory of executive functions (Miyake et al., 2000) posits that interrelated but separate components of cognitive functioning include cognitive flexibility, working memory, and inhibition as basic components. In detail, cognitive flexibility refers to the ability to change perspective or approach a problem with new rules or priorities. Working memory involves the ability to hold and mentally manipulate information. Finally, inhibition refers to the ability to stop an automatic or dominant response in order for a person to achieve a certain goal or delay the satisfaction of certain needs. Executive functions are most often assessed using behavioural tasks and rating scales. Behavioural tasks typically measure accuracy and reaction time. On the other hand, rating scales have been developed to improve the ecological validity of executive function measures in complex everyday situations. Using scales to assess executive function, the participant reports cognitive difficulties performing everyday tasks. The use of both behavioural tasks and rating scales as part of executive function assessment is recommended, as different methods have been shown to measure different aspects of executive function (Toplak et al., 2013).

Many studies suggest that depression is associated with impaired executive functions. For example, individuals with depressive symptoms performed poorly on executive function tasks such as working memory (Semkovska et al., 2019) and inhibition (Yitzhak et al., 2023) compared to controls. According to the Resource Allocation Model (RAM; Ellis & Moore, 1999), it is assumed that there is a fixed amount of resources available for information processing. According to this model, depression increases cognitive load and consumes resources that would otherwise be devoted to

a task. In other words, a higher level of depression would reduce a person's ability to focus attention on cognitive tasks, especially in the case of complex tasks. In addition, some researchers believe that executive functions can serve as a protective factor that aids coping with depressive symptoms; that is, they believe executive function training can reduce the likelihood of depressive symptoms (Beloe & Derakshan, 2020). Studies suggest that better executive functions most likely "protect" an individual from preoccupation with negative thoughts and low mood. However, studies in this research field are not entirely consistent. In particular, some studies show no differences between depressed and non-depressed individuals on measures of executive functions (Murphy et al., 2019), except in late adulthood (Rosselli et al., 2019). Working memory and, particularly, inhibition seem to be the most frequently studied executive functions in depression-related research. Specifically, inhibition is mentioned in 68%, and working memory is mentioned in 35% of research studies, which makes them the most represented in the literature on executive functions (Baggetta & Alexander, 2016). Therefore, it is necessary to determine their separate contribution to the mental health of young adults. Additionally, the link between executive functions and depressive symptoms is well-documented in clinical and older adult samples; however, the relation between these constructs within non-clinical samples of young adults is less recognized. Exploring factors associated with depression in young adults in Eastern Europe is particularly important, given that this population was shown to be at increased risk of serious mental health problems.

The Present Study

The aim of the present study was to explore the association of working memory and inhibition with symptoms of depression in a non-clinical sample of young adults in Croatia using a cross-sectional design. Given that some previous research has used only subjective or only objective measures of executive functions (i.e., Nikolin et al., 2021), the strength of our research design is the parallel use of a subjective self-report questionnaire, as well as objective behavioural tasks to measure specific executive functions and their separate relations to depressive symptoms. In line with the Resource Allocation Model of Ellis & Moore (1999), it was hypothesized that better working memory and inhibitory ability, as measured by self-report and

behavioural tasks, would be associated with lower levels of depressive symptoms in young adults.

Method

Data and Participants

The data were collected online during August and September 2023 using a snowball method via social networks, portals, e-mail lists, and by directly sharing the link with acquaintances. To achieve representativeness, the sample was weighted by demographic characteristics (i.e., age, socioeconomic status) and device (computer and mobile). After excluding participants with neurological disorders (i.e., epilepsy, traumatic brain injury), the final sample consisted of 290 young adults (225 females, 64 males, and one non-binary individual) aged 18-35 (M_{age} = 24.35 years, SD_{age} = 3.87 years). The participant's level of education was as follows: 0.3% had finished primary school, 35.4% had finished high school (between 11 and 12 years of schooling), 63.6% had a college degree, and 0.7% had a postgraduate degree. Moreover, 73.7% reported an average financial status. With regard to their employment status, 2% were high school pupils, 54.8% were students, 35.3% were employed and 7.9% unemployed. The research was ethically approved by the University Institutional Review Board (IRB). To be allowed to participate in the study, all participants provided written informed consent.

Measures

Executive functions – self-report

Adult Executive Function Inventory (ADEXI; Štelcar, 2021)

The Croatian version of the ADEXI (Štelcar, 2021) was used to measure executive functions in young adults. The inventory consists of 14 items measuring working memory difficulties (9 items; e.g., "I have difficulties with tasks or activities that involve several steps") and inhibition difficulties (5 items; e.g., "I have a tendency to do things without first thinking about what could happen"). The participants used a five-point Likert-type scale to assess how well each statement describes them (1 = "definitely not true," 2 = "not

true," 3 = "partially true," 4 = "true," and 5 = "definitely true"). The score is calculated as a linear combination of answers to the associated statements, with higher scores indicating greater difficulties in executive functions. In our sample of young adults, the internal reliability coefficients of the subscales were α = .87 (working memory difficulties) and α = .66 (inhibition difficulties).

Executive functions – behavioural tasks

We used the free software package OpenSesame (Mathôt & March, 2022) to construct computerized behavioural tasks for executive function (working memory and inhibition) assessment.

The Stroop task (Mead et al., 2002)

During the Stroop task, participants were presented with stimuli in the form of colour names coloured in different colours on the screen. The presented stimulus was either congruent or incongruent. When the stimulus was congruent, word colour and word name were matched (i.e., the word “red” was presented in red colour), whereas, in the case of incongruent stimuli, the word and the colour of the word were not matched (i.e., the word "red" was presented in yellow colour). The participants' task was to determine the colour of each displayed word by pressing the correct key on the keyboard (i.e., if the word "red" was displayed in green colour, they were to press the "g" key, which indicates the colour green). The success rate was determined by the average difference in reaction speed to congruent and incongruent stimuli (expressed in milliseconds). A smaller difference in reaction speed reflects the participant’s better ability to inhibit dominant responses. The test-retest reliability of the Stroop task in earlier samples of young adults was high, ranging from 0.78 to 0.92 (Vora et al., 2016).

The N-back task (version used by Miller et al., 2009)

The N-back task was used to assess participants' working memory ability. Participants were presented with a sequence of three sets of 25 letters. Each letter was presented separately, and the task was to assess whether the presented letter had already been shown in two previous presentations of the letter. When a letter was presented in a predetermined order, the participant's task was to press the corresponding key on the keyboard (i.e., participants are first shown the letter "F," then the letter "B,"

then again the letter "F"). When a letter was not presented in a predetermined order, the participant was required not to respond. The indicator of success was the percentage of correctly recognized displays, with higher scores representing better working memory ability. The test-retest reliability of this task in earlier samples was .85 (Soveri et al., 2016).

Depressive symptoms

Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001)

The Croatian version of the PHQ (Bolić, 2024) was used to measure depressive symptoms in young adults. For each of the nine statements in the questionnaire, participants estimated how often they experienced certain symptoms in the last two weeks (e.g.," Feeling tired or having little energy"). Participants used a four-point Likert-type scale (0 = "Not at all," 3 = "Nearly every day"). The total score was calculated as a linear combination, with higher scores indicating a severity of depressive symptoms. The internal reliability coefficient of this questionnaire in our study was α = .79.

Control variables

Participants' education level and financial status, as well as the level of resilience measured by the Connor-Davidson Resilience Scale 10-CD-RISC-10 (Campbell-Sills & Stein, 2007), were used as control variables in this research. For a detailed description of the control variables, see Bolić (2024).

Results

Preliminary Analysis

The Kolmogorov-Smirnov (KS) test, skew index (SI), and kurtosis index (KI) were used to check the normality of the distributions. The Kolmogorov-Smirnov test suggested the distribution of all variables significantly differed from normal ($p < .05$). However, noting that even one extreme result can cause the distribution to deviate significantly from normal, Kline (2011) explains that the distribution can still be considered normal if the absolute values of SI < 3.0 and KI < 10.0. In our study, the SI of

all variables ranged from .02 to 2.15, while the KI values ranged from .08 to 5.36.

Partial correlation analyses were used to explore research hypotheses. As is shown in Table 1, when controlling for education level, financial status, and level of resilience, young adults with more working memory difficulties (measured with self-report or behavioural tasks) had higher levels of depressive symptoms. However, inhibition was not related to depressive symptoms when assessed via self-report or behavioural task. Thus, inhibition will not be included in further analyses. As for the relationship between different methods for assessing executive functions (i.e., self-report and behavioural measures), the correlations ranged from non-significant to low.

Table 1

Descriptive Statistics, Normality Tests, and Partial Correlations

Variables		*M (SD)*	*KS*	*SI*	*KI*	*2*	*3*	*4*	*5*
1	ADEXI working memory	20.92 (6.29)	.09	0.60	0.08	**.44****	.04	.02	**.12***
2	ADEXI inhibition	13.44 (3.79)	.08	0.02	-0.34	-	.03	**.17****	.08
3	N-back working memory	77.53 (20.25)	.07	-2.15	5.36		-	**.13****	**-.14***
4	Stroop inhibition	139.60 (159.11)	.22	-0.44	-1.39			-	.01
5	PHQ-9	7.17 (4.16)	.13	0.47	-0.50				-

Note. KS – Kolmogorov-Smirnov test; SI – Skew Index; KI – Kurtosis Index.

$^{*}p < .05$. $^{**}p < .01$.

The role of working memory in explaining depressive symptoms

Following partial correlation analyses, we wanted to examine whether self-report measure or behavioural task of working memory will have

greater contribution in explaining depressive symptoms. Therefore, hierarchical regression analysis was conducted. The first block of predictors consisted of control variables. The second block included measures of working memory difficulties.

As can be seen in Table 2, after controlling for sociodemographic variables, self-reported working memory difficulties were not a significant predictor of depressive symptoms in our sample of young adults. However, the behavioural task of working memory (N-back task) explains 2% ($p < .05$) of the variance of depressive symptoms. In line with the results of the correlation analysis, a higher number of correct results on the N-back task, which indicates better working memory ability, independently predicts lower scores on the depression symptoms questionnaire.

Partially in line with our hypothesis, only the N-back task was a significant predictor of depressive symptoms. In other words, better working memory, as measured by a behavioural task, predicted lower levels of self-reported depressive symptoms in young adults.

Table 2

The Contribution of Working Memory to the Explanation of Depressive Symptoms

Variables	ß
1. step	
Education level	-.05
Financial status	-.07
Resilience	-.09
F	2.14
df_1	3
df_2	286
ΔR^2	.01
2. step	
Education level	-.05
Financial status	-.07
Resilience	-.09

	ADEXI working memory difficulities	.07
	N-back task	-.13*
F		3.04
df_1		5
df_2		284
ΔR^2		.02*

* $p < .05$.

Discussion

In a sample of young adults, we wanted to investigate the association of working memory and inhibitory ability with symptoms of depression. Specifically, we wanted to test whether better working memory and inhibition would be associated with lower levels of depressive symptoms.

This study extends earlier research on the link between executive functions and depressive symptoms by examining specific relations between working memory and inhibition and depressive symptoms in a non-clinical sample of young adults using parallel versions of self-report and behavioural tasks that measure executive functions. Consistent with the work of Ellis & Moore (1999), our study showed that executive functions may be associated with symptoms of depression in a sample of young adults. However, there were certain unexpected findings that we will elaborate in the following paragraphs.

Partially in line with our hypothesis and similar to previous research, people with difficulties in working memory, as measured by self-assessments (Hoorelbeke et al., 2022) and behavioural measures (Nikolin et al., 2021), reported more pronounced symptoms of depression. Many studies suggest that depressed individuals have dysfunctions in the activities of parts of the brain responsible for cognitive processes involved in working memory, such as selective attention, updating and information manipulation (Wang et al., 2021). These results are consistent with the Resource Allocation Model (Ellis & Moore, 1999), according to which depression increases cognitive load and reduces the ability to focus attention on cognitive tasks, especially in the case of complex tasks. Overall, in our study, it appeared that participants with higher levels of depressive symptoms had

working memory difficulties, as reflected in their poorer performance on the N-back task. Besides, they seemed to be aware of the difficulties working memory deficits cause in their daily lives, as reflected in their self-reports. Given that the research used neutral, non-emotional stimuli, it can be assumed that the general ability to process information, regardless of the type of stimulus, predicts symptoms of depression. In addition, there is evidence of a neurobiological basis for this association. Precisely, depressive symptoms are associated with structural abnormalities in the prefrontal lobe and posterior cortical regions (i.e., Walsh et al., 2007), which lead to less activation of these brain areas and lower levels of glutamate - the neurotransmitter associated with memory, cognition and emotion regulation (i.e., Pehrson & Sanchez, 2014). On the other hand, unlike the research of Hoorelbeke et al. (2022), in our study the ADEXI working memory difficulties subscale was not a significant predictor of depressive symptoms. This can be attributed to the somewhat low correlation between the ADEXI working memory difficulties subscale and depressive symptoms subscale, compared to the working memory behavioural task.

It is worth pointing out that the low correlations between executive functions and depressive symptoms are not surprising given that our study was conducted on a non-clinical sample of young adults, compared to stronger correlations in the research conducted on the same age groups with diagnosed clinical disorders, such as major depressive disorder (i.e., Liu et al., 2019). Additionally, similar to the results of our study, results from previous research (e.g., Toplak et al., 2013) show non-significant to low correlations between behavioural measures of executive function and executive function rating scales. This indicates that behavioural tasks and executive function rating scales measure different aspects of the same construct. Toplak et al. (2013) believe that different executive function measures capture different levels of cognitive analysis. Specifically, behavioural tasks are thought to provide information about the efficiency of information processing, while rating scales provide an understanding of an individual's success in achieving certain goals. In addition, Toplak et al. (2013) state that differences in the understanding of behavioural tasks and rating scales also arise from the assumptions of measurement theory in psychology. Namely, in psychometrics, there are two types of performance, that is, reactions expected from participants in a certain situation, namely:

(1) typical or usual and (2) optimal or maximal performance of a certain activity or task. In rating scales, the participant's reaction is not completely defined by the examiner's instructions, and the participant is not expected to provide their best possible performance. Instead, the participant has the freedom to interpret the situation in their own way. Conversely, in behavioural tasks, there is a clearly defined situation in which participants are expected to perform optimally, with their task being to achieve as much success as possible. Given the above, future research should focus on a more detailed exploration of various behavioural correlates of rating scales and behavioural tasks for assessing executive functions.

Contrary to expectations, no association was found between any of the two measures of inhibition (i.e., subjective and objective) and depressive symptoms. Such findings are not in accordance with many previous studies (i.e., Yitzhak et al., 2023). However, some researchers (e.g., Rosselli et al., 2019) argue that inconsistent findings on the executive functions-depression link can be found in a sample of adolescents, as deficits in executive functions in individuals with depressive disorders are more common in late adulthood. Moreover, the lack of consistency across studies may be explained by participants' level of depressive symptoms. That is, the effect size of this association increases as the level of depressive symptoms increases (Liu et al., 2019), and in our study, participants reported mild depressive symptoms on average. In addition, a longitudinal study by Yitzhak et al. (2023) suggested there are intra-individual fluctuations in inhibitory capacity measured by the Go-NoGo behavioural task; specifically, inhibitory capacity may fluctuate at a daily level. In their study, poor performance on the inhibition-related task was not observed during the initial measurement but only after five consecutive days of measurement. Therefore, it is possible that our study did not find a statistically significant association between inhibition and depressive symptoms due to intra-individual daily fluctuations in inhibition. Also, it would be useful to test the relationship between inhibition and depressive symptoms using another behavioural task measuring inhibition ability, such as the Go-NoGo task. The participant's task is to react or perform a motor action during the *go* state and to resist the impulsive urge to do so during the *no-go* state. A higher proportion of incorrect responses in the task is considered a direct measure of inhibition-related difficulties. Therefore, this type of task is recommended for future research.

All the above points to the importance of studying each executive function individually and determining their separate contribution when it comes to different difficulties.

Our study points to several factors that should be considered in future research. For example, the nature of its implementation. Given that the study was conducted online, it was not possible to control for some systematic variable factors such as previous exposure to similar behavioural tasks. Given that executive function tasks cannot be paused once the participant has begun to solve them, various distractions may have led to poorer performance. Furthermore, due to the age of the participants, it is possible that at least some of them play video games, and people who regularly play video games perform better on many measures of executive functions (i.e., Alho et al., 2022); thus, future research would benefit from data on the frequency of playing video games in daily life. Furthermore, self-assessment measures are prone to subjectivity; participants may have provided socially desirable responses, and their responses may have been influenced by their current mood, their understanding of the research concepts, and their ability to accurately rate their difficulties. In addition, due to the lower reliability of the ADEXI inhibition difficulties subscale, results related to this subscale should be interpreted with caution. However, it should be noted that, according to some authors (e.g., Taber, 2017), an acceptable level of reliability may be lower than $\alpha = .70$, depending on the constructs being measured. Earlier interpretations by Gardner (1995) emphasized that a rating scale must be unidimensional in order to produce interpretable results. It is possible that other self-report measures, such as the Behavior Rating Inventory of Executive Function-Adult Version (Roth et al., 2005), which has excellent psychometric properties, may provide better insight into the relationship between executive functions and depression in youth. Given the low percentage of explained variance of the criterion variable of depressive symptoms, future research may include tasks that measure cognitive flexibility and some higher-order executive functions such as reasoning, problem-solving, and planning. For example, future research could use the Wisconsin Card Sorting Task (Stuss et al., 2000), in which the participant has to adapt their behaviour or thoughts to the new demands of the situation, as well as the Tower of London (Kaller et al., 2012), which was designed to measure planning ability. The inclusion of different measures of executive

functions and related cognitive abilities, such as cognitive reserve (e.g., Volarov et al., 2020), could provide more detailed information on the association of specific cognitive functions with depressive symptoms. Moreover, future research could include other variables related to depressive symptoms, such as personality traits, emotion regulation strategies, sleep habits, and physical activity (e.g., Schuch et al., 2017). Additionally, our sample predominantly consisted of females, so caution is needed when interpreting the results, given that gender is a well-known covariate of the studied variables (i.e., Gaillard et al., 2021; Hyde & Mezulis, 2020). Although the level of education was included as a control variable, since the number of years of education is associated with better cognitive functions throughout adulthood (Lövdén et al., 2020), it is important to note that our sample predominately consisted of young, highly educated people Also, higher education is considered a protective factor in the development of mental health problems such as depression in youth (Bauldry, 2015). Another limitation of the present study is its correlational design, making it impossible to establish a causal relationship between the variables of interest. The strengths of the present study are worth noting, too. This research focuses on the relatively unexplored associations between specific dimensions of executive functions and depression in young adults. Moreover, the present study used both objective (i.e., computerized behavioural tasks) and subjective (i.e., self-reported) measures of executive functions. By using different methods for assessing executive functions, a more valid prediction of depressive symptoms in early adulthood is possible. Such a practice takes into account that some high correlations may be attributable to common method variance. Furthermore, this research places executive function in the context of common problems of modern life. A large number of young adults have experienced at least one major depressive episode during their lifetime, so it is important to discover mechanisms that can help them cope with these mental health problems. The conducted research showed that working memory has a relevant role in explaining the symptoms of depression. The results can help in designing strategies aimed at reducing the symptoms of depression through executive functions, especially working memory training for young adults, both computerized and noncomputerized. Moreover, as pointed out by Novick et al. (2020), the way in which a certain activity is performed as part of the executive function training, the characteristics of the

mentor or trainer, as well as the perception of the importance and relevance of the activity for the participant will likely have a more significant effect on the participant than the activity itself. Together, the results of the conducted study can serve as a basis for further research in different age groups and in the context of different life challenges.

Conflict of interest

We have no conflicts of interest to disclose.

Data availability statement

Data used in this paper are available upon a reasonable request.

References

Alho, K., Moisala, M., & Salmela-Aro, K. (2022). Effects of media multitasking and video gaming on cognitive functions and their neural bases in adolescents and young adults. *European Psychologist*, *27*(2), 131–140. https://doi.org/10.1027/1016-9040/a000477

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. https://doi.org/10.1176/appi.books.9780890425596

Arias-de la Torre, J., Vilagut, G., Ronaldson, A., Bakolis, I., Dregan, A., Martín, V., Martinez-Alés, G., Molina, A. J., Serrano-Blanco, A., Valderas, J. M., Viana, M. C., Alonso, J. (2023). Prevalence and variability of depressive symptoms in Europe: update using representative data from the second and third waves of the European Health Interview Survey (EHIS-2 and EHIS-3). *Lancet Public Health*, *8*(11), e889-e898. https://doi.org/10.1016/S2468-2667(23)00220-7

Baggetta, P., & Alexander, P. A. (2016). Conceptualization and operationalization of executive function. *Mind, Brain, and Education*, *10*(1), 10–33. https://doi.org/10.1111/mbe.12100

Bauldry, S. (2015). Variation in the protective effect of higher education against depression. *Society and Mental Health*, *5*(2), 145–161. https://doi.org/10.1177/2156869314564399

Beloe, P., & Derakshan, N. (2020). Adaptive working memory training can reduce anxiety and depression vulnerability in adolescents. *Developmental Science*, *23*(4) e12831. https://doi.org/10.1111/desc.12831

Bolić, T. (2024). Odnos izvršnih funkcija i otpornosti s depresivnim simptomima kod mladih odraslih osoba [Neobjavljeni diplomski rad, Sveučilište Josip Juraj Strossmayera u Osijeku]. Filozofski fakultet.

Campbell-Sills, L. & Stein, M. B. (2007). Psychometric analysis and refinement of the Connor-Davidson Resilience Scale (CD-RISC): Validation of a 10-item measure of resilience. *Journal of Traumatic Stress*, *20*(6), 1019–1028. https://doi.org/10.1002/jts.20271

Ellis, H. C. & Moore, B. A. (1999). Mood and memory. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 193–210). John Wiley & Sons Ltd. https://doi.org/10.1002/0470013494.ch10

Frewen, P., Zhu, J., & Lanius, R. (2019). Lifetime traumatic stressors and adverse childhood experiences uniquely predict concurrent PTSD, complex PTSD, and dissociative subtype of PTSD symptoms whereas recent adult non-traumatic stressors do not: results from an online survey study. *European Journal of Psychotraumatology*, *10*(1), 1606625. https://doi.org/10.1080/20008198.2019.1606625

Gaillard, A., Fehring, D. J., & Rossell, S. L. (2021). A systematic review and meta-analysis of behavioural sex differences in executive control. *European Journal of Neuroscience*, *53*(2), 519-542. https://doi.org/10.1111/ejn.14946

Gardner, P. L. (1995). Measuring attitudes to science: Unidimensionality and internal consistency revisited. *Research in Science Education*, *25*(3), 283–289. https://doi.org/10.1007/bf02357402

Goodwin, R. D., Dierker, L., Wu, M., Galea, S., Hoven, C. W., & Weinberger, A. H. (2022). Trends in U.S. depression prevalence from 2015 to 2020: The widening treatment gap. *American Journal of Preventive Medicine*, *63*(5), 726–733. https://doi.org/10.1016/j.amepre.2022.05.014

Hoorelbeke, K., Zwalmen, Y. V., Hagen, B. I., Stubberud, J., Koster, E. H.W. (2022). Connecting residual depressive symptoms to self-reported executive functioning: A network analytical approach, *Journal of Psychiatric Research*, *155*, 75-84, https://doi.org/10.1016/j.jpsychires.2022.08.007

Hyde, J. S., & Mezulis, A. H. (2020). Gender differences in depression. *Harvard Review of Psychiatry*, *28*(1), 4–13. https://doi.org/10.1097/hrp.0000000000000230

Kaller, C. P., Unterrainer, J. M., & Stahl, C. (2012). Assessing planning ability with the Tower of London task: Psychometric properties of a structurally balanced problem set. *Psychological Assessment*, *24*(1), 46–53. https://doi.org/10.1037/a0025174

Kline, R. B. (2011). *Principles and practice of structural equation modeling (3rd ed.)*. Guilford Press.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Liu, T., Zhong, S., Wang, B., Liao, X., Lai, S., & Jia, Y. (2019). Similar profiles of cognitive domain deficits between medication-naive patients with bipolar II depression and those with major depressive disorder. *Journal of Affective Disorders*, *243*, 55–61. https://doi.org/10.1016/j.jad.2018.05.040

Lövdén, M., Fratiglioni, L., Glymour, M. M., Lindenberger, U., & Tucker-Drob, E. M. (2020). Education and cognitive functioning across the life span. *Psychological Science in the Public Interest*, *21*(1), 6–41. https://doi.org/10.1177/1529100620920576

Mathôt, S. & March, J. (2022). Conducting linguistic experiments online with OpenSesame and OSWeb. *Language Learning*, *72*, 1017-1048. https://doi.org/10.1111/lang.12509

Mead, L. A., Mayer, A. R., Bobholz, J. A., Woodley, S. J., Cunningham, J. M., Hammeke, T. A., & Rao, S. M. (2002). Neural basis of the Stroop interference task: response competition or selective attention? *Journal of the International Neuropsychological Society*, *8*(6), 735-742. http://doi.org/10.1017/s1355617702860015

Miller, K. M., Price, C. C., Okun, M. S., Montijo, H., & Bowers, D. (2009). Is the N-Back task a valid neuropsychological measure for assessing working memory? *Archives of Clinical Neuropsychology*, *24*(7), 711–717. https://doi.org/10.1093/arclin/acp063

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Murphy, O. W., Hoy, K. E., Wong, D., Bailey, N. W., Fitzgerald, P. B., & Segrave, R. A. (2019). Individuals with depression display abnormal modulation of neural oscillatory activity during working memory encoding and maintenance.

Biological Psychology, *148*, 107766. https://doi.org/10.1016/j.biopsycho.2019.107766

Nikolin, S., Tan, Y. Y., Schwaab, A., Moffa, A., Loo, C. K.; & Martin, D. (2021). An investigation of working memory deficits in depression using the n-back task: A systematic review and meta-analysis. *Journal of Affective Disorders*, *284*, 1–8. https://doi.org/10.1016/j.jad.2021.01.084

Novick, J. M., Bunting, M. F., Dougherty, M. R., & Engle, R. W. (Eds.). (2020). *Cognitive and working memory training: Perspectives from psychology, neuroscience, and human development*. Oxford University Press.

Pehrson, A.L. & Sanchez, C. (2014). Serotonergic modulation of glutamate neurotransmission as a strategy for treating depression and cognitive dysfunction. *CNS Spectrums*, *19*(2), 121-133. https://doi.org/10.1017/S1092852913000540

Roth, R. M., Isquith, P. K., & Gioia, G. A. (2005). *Behavior Rating Inventory of Executive Function®--Adult Version (BRIEF-A)* [Database record]. APA PsycTests. https://doi.org/10.1037/t86244-000

Rosselli, M., Lang, M.K., & Arruda, F. (2019). Executive dysfunction in depressive disorders. In A. Ardila, A., S. Fatima & M. Rosselli (Eds.), *Dysexecutive syndromes: Clinical and experimental perspectives* (pp. 241-259). Springer Nature Switzerland AG.

Schuch, F., Vancampfort, D., Firth, J., Rosenbaum, S., Ward, P., Reichert, T., Carvalho Bagatini, N., Bgeginski, R., & Stubbs, B. (2017). Physical activity and sedentary behavior in people with major depressive disorder: A systematic review and meta-analysis. *Journal of Affective Disorders*, *210*, 139–150. https://doi.org/10.1016/j.jad.2016.10.050

Semkovska, M., Quinlivan, L., O'Grady, T., Johnson, R., Collins, A., O'Connor, J., Knittle, H., Ahern, E., Gload, T. (2019). Cognitive function following a major depressive episode: a systematic review and meta-analysis. *The Lancet Psychiatry*, *6*(10), 851-861. https://doi.org/10.1016/S2215-0366(19)30291-3

Soveri, A., Lehtonen, M., Karlsson, L. C., Lukasik, K., Antfolk, J., & Laine, M. (2016). Test–retest reliability of five frequently used executive tasks in healthy adults. *Applied Neuropsychology: Adult*, *25*(2), 155–165. https://doi.org/10.1080/23279095.2016.1263795

Stuss, D. T., Levine, B., Alexander, M. P., Hong, J., Palumbo, C., Hamer, L., Murphy, K. J., & Izukawa, D. (2000). Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: effects of lesion

location and test structure on separable cognitive processes. *Neuropsychologia*, *38*(4), 388-402. https://doi.org/10.1016/S0028-3932(99)00093-7

Štelcar, M. (2021). *Odnos izvršnih funkcija i stilova rukovođenja* [Diplomski rad, Sveučilište Josip Juraj Strossmayera u Osijeku]. Filozofski fakultet. https://urn.nsk.hr/urn:nbn:hr:142:250644

Taber, K. S. (2017). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry*, *54*(2), 131–143. https://doi.org/10.1111/jcpp.12001

Volarov, M., Miladinović, P., Gligorijević, A., Poznanović, M., Obradović, J., Nikolić, N., & Bugarski-Ignjatović, V. (2020). Serbian cognitive reserve index questionnaire: Adaptation and validation. In G. Nedović, & F. Eminović (Eds.), *Approaches and Models in Special Education and Rehabilitation* (pp. 11-24). University of Belgrade – Faculty of Special Education and Rehabilitation. https://www.icf.fasper.bg.ac.rs/zbornici/20200825_1-Thematic-Collection-2020.pdf

Vora, J.P., Varghese, R., Weisenbach, S.L., Bhatt, T. (2016). Test-retest reliability and validity of a custom-designed computerized neuropsychological cognitive test battery in young healthy adults. *Journal of Psychology and Cognition*, *1*(1), 11-19. https://pubmed.ncbi.nlm.nih.gov/28286883/

Walsh, N. D., Williams, S. C. R., Brammer, M. J., Bullmore, E. T., Kim, J., Suckling, J., Mitterschiffthaler, M. T., Cleare, A. J., Pich, E. M., Mehta, M. A., & Fu, C. H. Y. (2007). A longitudinal functional magnetic resonance imaging study of verbal working memory in depression after antidepressant therapy. *Biological Psychiatry*, *62*(11), 1236–1243. https://doi.org/10.1016/j.biopsych.2006.12.022

Wang, X., Cheng, B., Roberts, N., Wang, S., Luo, Y., Tian, F., & Yue, S. (2021). Shared and distinct brain fMRI response during performance of working memory tasks in adult patients with schizophrenia and major depressive disorder. *Human Brain Mapping*, *42*(16), 5458–5476. https://doi.org/10.1002/hbm.25618

Yitzhak, N., Shimony, O., Oved, N., Bonne, O., & Nahum, M. (2023). Less inhibited and more depressed? The puzzling association between mood, inhibitory control and depressive symptoms. *Comprehensive Psychiatry*, *12*, Article 152386. https://doi.org/10.1016/j.comppsych.2023.152386

Primenjena psihologija
Vol. 18, No. 2, pp. 227–254, 2025

Research Article

Psychometric evaluation of an Indonesian version of the Brief Self-Control Scale: Item calibration using polytomous Rasch models

Diana Mutiah [1] ✉, Bobby Suwandi [2], and Muhammad Dwirifqi Kharisma Putra [3]

[1] *Faculty of Psychology, Universitas Islam Negeri Syarif Hidayatullah Jakarta, Indonesia*
[2] *Faculty of Psychology, Universitas Pancasila, Indonesia*
[3] *Faculty of Psychology, Universitas Gadjah Mada, Indonesia*

ABSTRACT

This study aimed to evaluate the psychometric properties of the 10-item Indonesian version of the Brief Self-Control Scale (BSCS). It used the polytomous Rasch model, which enables more detailed analysis, including differential item functioning (DIF) analysis. The participants in this study were 1001 Indonesian high school students. We found that the partial credit model (PCM) was a better fit than the rating scale model. Furthermore, the unidimensionality, local independence, and monotonicity assumptions of the PCM were valid for the BSCS. Q5 was the only item that did not fit the PCM. The step parameters of the BSCS functioned well, with values ranging from low to high, as expected, for all items, indicating monotonicity. Person separation reliability was 0.71, indicating that the BSCS has good internal consistency. The DIF analysis showed that item Q5 functioned differently across genders. In general, the remaining nine items of the BSCS have good psychometric properties for measuring self-control.

Keywords: BSCS, calibration, self-control, polytomous Rasch model, validation

UDK: 159.938.3
DOI: 10.19090/pp.v18i2.2559
Received: 11.06.2024.
Revised: 01.10.2024.
Accepted: 19.11.2024.

✉ Corresponding author's email: diana.mutiah@uinjkt.ac.id

Introduction

Self-control is 'the ability to override or change one's inner responses, as well as to interrupt undesired behavioral tendencies and refrain from acting on them' (Tangney et al., 2004). Research on self-control has grown rapidly in the last decade, especially in relation to emerging constructs in psychology, such as the relationship between self-control and self-discipline (Hagger et al., 2021), grit (Fekih-Romdhane et al., 2022), and loneliness (Stavrova et al., 2022). The construct of self-control has attracted substantial attention from psychologists working within a variety of theoretical and methodological frameworks (Duckworth & Kern, 2011; Gillebaart, 2018).

Various studies have used the theoretical basis provided by Tangney et al.'s (2004) definition to propose components of self-control. Maloney et al. (2012) suggested that impulsivity and restraint are aspects of self-control, whereas de Ridder et al. (2011) stated that inhibitory self-control and initiatory self-control are constituents of self-control. However, according to Tangney et al.'s (2004) initial concept, self-control is a unidimensional construct (see Manapat et al., 2021), meaning that it can be expressed as a single continuum, from less or low self-control to high self-control.

In the Indonesian context, a recent study found that self-control is associated with higher pro-environmental behaviour of Indonesians (Zwagery et al. 2023). In a sample of high school students from Indonesia, more than 80% had moderate levels of self-control (Qonita & Herdi, 2023). In addition, high self-control was found to be associated with lower stress in a sample of nurses from Indonesia (Paramitha & Ariani, 2024). These findings indicate that self-control has been widely studied in Indonesia in recent years.

From a methodological standpoint, self-control measures vary in operational definition and procedure, from simple questionnaires to complex scenarios where individuals must choose whether or not to act (Milyavskaya et al., 2019; Pilcher et al., 2023). Among these various procedures and scales, a tool that is widely used to measure self-control is the Brief Self-Control Scale (BSCS; Tangney et al., 2004). The BSCS was developed from the 36-item Self-Control Scale (SCS) that has a five-dimensional factor structure, although it is scored as a unidimensional one-factor model

(Manapat et al., 2021). The BSCS is widely used because of its Likert-type response format, which is easier to apply compared to other formats (i.e., executive function tasks or delay of gratification tasks; for more details, please see Duckworth & Kern, 2011). In addition, shorter scales, such as the BSCS, are particularly useful in real-world settings, where time and resources are often limited (Pechorro et al., 2021).

The BSCS has been adapted into various languages, for example, Arabic (Fekih-Romdhane et al., 2022), Chinese (Fung et al., 2020), Dutch (Kupper et al., 2020), French (Brevers et al., 2017), German (Bertrams & Dickhäuser, 2009), Greek (Papanikolopoulos et al., 2022), Indonesian (Arifin & Milla, 2020; Zwagery et al., 2023), Italian (Chiesi et al., 2020), Japanese (Ozaki et al., 2016), Persian (Asgarian et al., 2020), Portuguese (Pechorro et al., 2021), Spanish (García-Castro et al., 2024), and Turkish (Nebioglu et al., 2012). Given the broad demographic range of application of the BSCS, population-representative studies have found the instrument to show good stability (Cobb-Clark et al., 2023).

However, since its initial development, the factor structure or dimensionality of the BSCS has changed across studies (Manapat et al., 2021; Papanikolopoulos et al., 2022). The original scale, the 36-item SCS, from which the BSCS is derived, consisted of five dimensions but used a single score with a unidimensional factor model (Manapat et al., 2021). The BSCS has been applied as a two-factor model (e.g., de Ridder et al., 2011; Ferrari et al., 2009; Maloney et al., 2012), while the original version was a one-factor model (e.g., Tangney et al., 2004). Consequently, previous studies (e.g., Hagger et al., 2021; Papanikolopoulos et al., 2022) have tested various measurement models named after the researchers who developed them, such as the ‘Maloney model’, ‘de Ridder model’, etc.

Compared to the 13 items of the original version of the BSCS (Tangney et al., 2004), subsequent versions had different numbers of items, e.g. 10 items (de Ridder et al., 2011) and 8 items (Maloney et al., 2012). Arifin and Milla (2020) based the Indonesian version of the BSCS on de Ridder et al.’s (2011) version consisting of 10 items with two dimensions - inhibition and initiation. This instrument was also used by other studies from Indonesia (Paramitha & Ariani, 2024). Conversely, Zwagery et al. (2023) used Ferrari et al.’s (2009) version with 13 items that measure two dimensions (self-

discipline and impulse control). Other Indonesian studies (Qonita & Herdi, 2023) used Tangney et al.'s (2004) 13-item version of BSCS as a unidimensional measure, but they removed three invalid items, leaving 10 items. Thus, four studies from Indonesia used a unidimensional measure with single scores.

To develop the initial version of the BSCS, Tangney et al. (2004) applied a classical test theory (CTT) approach. From a methodological perspective, CTT has many shortcomings, one of which is that the results of item analysis are highly dependent on the sample used (Andrich & Marais, 2019). The Indonesian version of the BSCS (Arifin & Milla, 2020; Paramitha & Ariani, 2024; Qonita & Herdi, 2023; Zwagery et al., 2023) has been validated using CTT and confirmatory factor analysis (CFA); however, as CFA is a congeneric model in CTT, the analysis has limitations such as the estimation of only one standard error of measurement for all respondents (Rusch et al., 2017).

In general, because of the limitations of CTT, various studies have evaluated the psychometric properties of instruments using modern test theory, i.e. item response theory (IRT), the Rasch model, and item factor analysis, which can mathematically overcome the limitations of CTT (Rusch et al., 2017). In the case of the BSCS, IRT (e.g. Manapat et al., 2021) and Rasch models (e.g. Chen et al., 2022) have provided more detailed item-analysis information compared to CTT. However, our literature review did not find any studies from Indonesia taking this approach.

Importantly, the application of the Rasch model to the Indonesian version of the BSCS would greatly assist non-specialist researchers in using this scale because the model provides a raw-score-to-logit conversion table (Saggino et al., 2020). In addition, by applying the Rasch model, measurement invariance or differential item functioning (DIF) in the Indonesian version of the BSCS instrument can be tested, as has been done with BSCS instruments from other countries. The BSCS has been found to be invariant across genders (Papanikolopoulos et al., 2022) and countries (Hagger et al., 2021). However, other studies have focused on testing gender differences in self-control measurement (Gibson et al., 2010; Jo & Bouffard, 2014). Although both studies used instruments other than BSCS, their results

indicate that DIF testing should be performed across genders for Indonesian samples in terms of self-control as a construct.

Therefore, this study aimed to evaluate the psychometric properties of the Indonesian BSCS instrument developed by Arifin and Milla (2020). Psychometric property testing was performed using the Rasch model, including a comparison of the partial credit model (PCM) and rating scale model (RSM) and testing of the assumptions. We believe that our study is the first to test for measurement invariance or DIF of the Indonesian BSCS.

Method

Participants

We used empirical data from 1001 respondents aged 12–19 years (*M* = 15.07 years, *SD* = 1.771 years); 387 females and 614 males participated in this study. A non-probability sampling method (i.e. quota sampling with a target of 1000 respondents) was used; data were collected over 4.5 months until the quota was reached. A Google form was sent to teachers, who then distributed it to their students. All participants received written information about the aim and procedures. They were also informed that participation was strictly voluntary and could be discontinued at any time without explanation. The information was attached to the questionnaire, and the participants gave their consent by completing the questionnaire.

The criteria used to determine the sample size were based on rule-of-thumb (Tennant & Küçükdeveci, 2023), according to which the minimum sample size for Rasch modelling was 250-500 respondents. It should be noted that this study was part of a larger project for determining pornography addiction among high school students; the BSCS was one of several instruments administered in this project.

Instruments

The Indonesian version of the Brief Self-Control Scale

The instrument used in this study is an Indonesian translation and adaptation by Arifin and Milla (2020) of Tangney et al.'s (2004) BSCS. The scale consists of 10 items (de Ridder et al., 2011), in contrast to the 13 original items. The response scale is a Likert-type scale with five response

options, ranging from *strongly disagree* (1) to *strongly agree* (5). In this study, we treated the Likert scale response options as ordered categorical data. Descriptive statistics of the items in our study are provided in Supplementary Materials.

The Rasch model

The Rasch model (Rasch, 1960) is a measurement theory developed by the Danish mathematician Georg Rasch. Mathematically, the Rasch model is simple, but from the perspective of measurement philosophy, it is profound (Mair, 2018). The Rasch model postulates that an individual's opportunity to correctly answer an item is determined by the interaction between two parameters - the item location and the person's trait level (Wu et al., 2016). The Rasch model uses a logit scale to present item difficulty parameters and individual trait level (or ability) parameters (Andrich & Marais, 2019).

Initially, the Rasch model was used only for analysing dichotomous data (e.g., 1 = 'true', 0 = 'false'). However, the Rasch model can also be used to analyse polytomous data, e.g. the Likert scale (Mair, 2018). The family of Rasch models developed to handle polytomous data includes the RSM (Andrich, 1978) and the PCM (Masters, 1982), both called polytomous Rasch models (PRMs; e.g., Andrich, 2013). In PRMs, one of the advantages of PCM parameterisation, compared to RSM parameterisation, is that it allows each item to have a different number of response categories (Andrich & Marais, 2019). Even when the instruments have the same number of response categories, the PCM provides information on the step parameter structure for each item. Then, the 'disordering of step parameters' of one or more items can be identified; on the other hand, it cannot be detected if the RSM is chosen without performing PCM analysis first (Wu et al., 2016).

The basic assumptions of the Rasch model, i.e. the unidimensionality and local independence assumptions, must be met (Andrich & Marais, 2019). The unidimensionality assumption postulates that all items in the measuring instrument measure a single latent variable. The local independence assumption assumes that an individual's response to one item should not influence their answer to another item (Mair, 2018). However, other assumptions, such as monotonicity (i.e., an assumption that

the probability of a positive response to an item [or, in the case of polytomous items, the transition from one response category to the next] should increase with underlying trait [e.g., self-control] levels) should also be tested (Tennant & Küçükdeveci, 2023).

In this study, the unidimensionality assumption was tested using the principal component analysis of residuals (PCAR; Smith, 2002). The local independence assumption was tested using the raw residual correlations between all pairs of items which is called Q3 statistics (Yen, 1984). Monotonicity was assessed by inspecting item threshold (step) patterns, which are expected by the model to monotonically increase from low to high across the continuum with no disordering (Tennant & Küçükdeveci, 2023).

Data analysis strategy

Because PRM has two parameterisations, rating scale parameterisation (RSM) and partial credit parameterisation (PCM), we performed both analyses and compared the global-fit statistics in the first phase. Models with better fit statistics were chosen. Mathematically, the comparisons between PCM and RSM are valid because the models are nested (Linacre, 2021). In the second phase, we tested the unidimensionality and local independence of the BSCS using the chosen model. In the third phase, we examined item fit statistics, including step parameters, to check for monotonicity. In the fourth phase, person-item maps or Wright Maps were reported. In the fifth phase, person reliability and item reliability were reported. In the sixth phase, the test information function was inspected. The seventh and last phase was DIF analysis. All phases were implemented in the Winsteps 5.1.4 program (Linacre, 2021) using unconditional or joint maximum likelihood estimation methods. We used the Winsteps-integrated 'wrightMap' package in R to create graphs (Irribarra & Freund, 2024).

Results

Global-fit statistics and model comparison

In the first phase, the two models, PCM and RSM, were analysed with two separate calibrations and the global-fit statistics of the models were compared. The global-fit statistics used were log-likelihood chi-squared (χ^2),

Akaike information criterion (AIC), and Bayesian information criterion (BIC), with a lower value indicating a better fitting model. Statistics based on residuals, root mean square residuals (RMSR), were also reported, with lower values indicating a better fitting model. Finally, the root mean square error (RMSE) was also compared, with a lower value indicating a better fitting model (Linacre, 2021; Welter et al., 2024). All global fit statistics and other statistics, such as RMSR and RMSE, indicated that the data for the Indonesian BSCS fit the PCM better than the RSM (for the model comparison results, see Table 1). The model comparison statistics AIC and BIC indicated that PCM fits the data better than RSM. Additionally, RMSR and RMSE were lower for PCM. Based on these findings, the model or parameterisation reported next is the PCM.

Table 1

Fit Statistics of the Rating Scale Model (RSM) and the Partial Credit Model (PCM)

Fit statistics	Model	
	RSM	PCM
Log-likelihood χ^2	20049.803	19442.345
AIC	22073.803	21520.345
BIC	29370.667	29011.889
RMSR	0.674	0.661
RMSE	0.520	0.516

Unidimensionality and local independence

In using PCAR, unidimensionality is achieved based on two criteria: first, the raw variance explained by measures should be greater than 40% (Holster & Lake, 2016); second, the unexplained variance in the 1st contrast should not exceed 2.0 (Smith, 2002). The results of the PCAR of the BSCS showed that 40.4% of the raw variance was explained by measures. The unexplained variance in the 1st contrast was 1.59 (Table 2).

Table 2

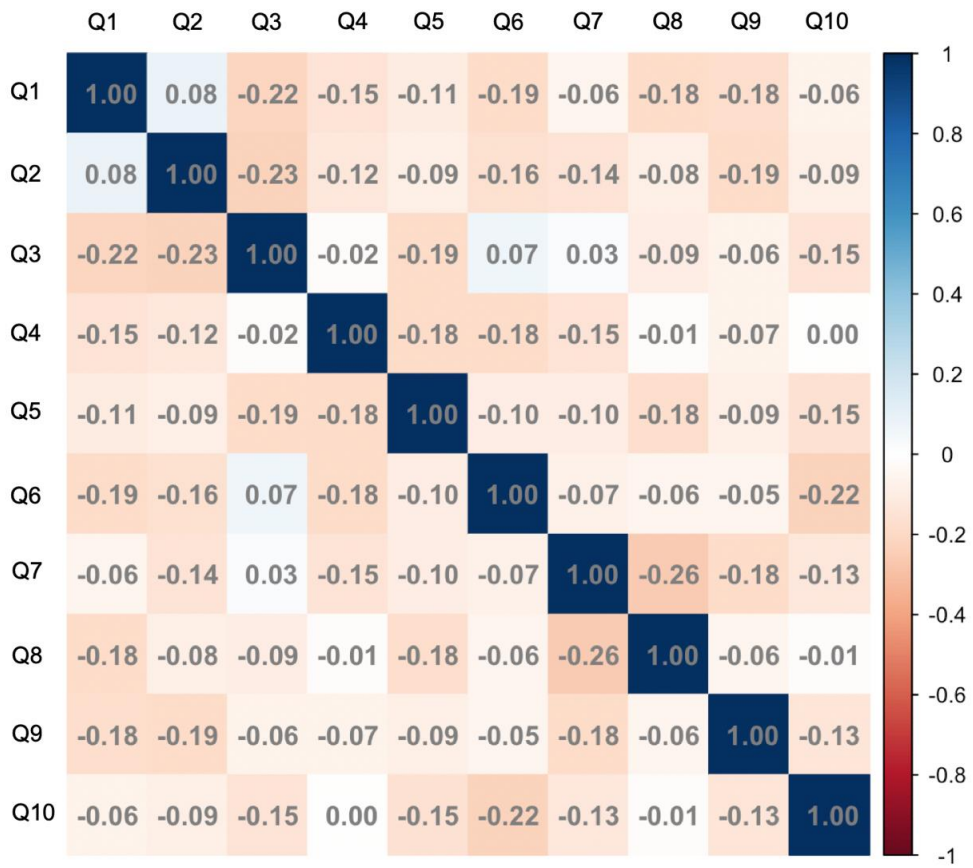
Results of the Principal Component Analysis of Residuals (PCAR)

PCAR statistics	Eigenvalue	Proportion of variance from observed data
Total raw variance	16.791	100%
Raw variance explained by measures	6.791	40.4%
Raw unexplained variance (total)	10.001	59.6%
Unexplained variance in the 1st contrast	1.590	9.5%

These results indicate that the unidimensionality assumption was met for the BSCS. In other words, the BSCS instrument was empirically found to measure one construct: self-control. Furthermore, after PCAR, the residual correlation matrix was inspected to check the assumption of local independence (see Figure 1).

Figure 1

Raw Residual Correlation Matrix (Q3 Statistics)



The results of the Q3 statistical test showed that the majority of items had a negative raw residual correlation (Figure 1). More specifically, there was no positive raw residual correlation with a value greater than 0.25 (DeMars, 2010), indicating that there was no substantial local dependency between the items. In addition, although items Q7 and Q8 had a residual correlation of -0.26, given the direction is negative, this is negligible. In other words, these results suggested that the assumption of local independence of the BSCS was met.

Item measure and fit statistics

After confirming that the assumptions of unidimensionality and local independence were valid for the BSCS, item parameter estimates and fit statistics were inspected (see Table 3). An item was deemed to fit the PCM if Infit and Outfit mean squares (MNSQs) were in the range of 0.5 to 1.5, with a point measure (PTMEA) correlation > 0.30 (Boone, 2020). Table 3 shows the items ordered from the most difficult to the easiest to endorse. The item location rangedbetween -0.971 and 1.603 logits. The more negative the difficulty level, the easier it was to obtain a higher score (a score of 5) on an item and vice versa. The easiest item on the BSCS scale was Q10, '*Saya menolak hal-hal yang buruk untuk diri saya*' [I refuse things that are bad for me], with a difficulty level of -0.971 logit. The most difficult item was Q5, '*Hal yang menyenangkan dan bersenang-senang kadang menahan saya untuk menyelesaikan pekerjaan*' [Pleasure and fun sometimes keep me from getting work done], with a difficulty level of 1.603 logits.

Table 3

Item Parameters, Fit Statistics, and Thresholds for All Items

Item	Measure	Infit MNSQ	Outfit MNSQ	PTMEA Corr.	Step 1	Step 2	Step 3	Step 4
Q5	1.603	1.68	1.67	-0.04	-2.68	0.08	3.12	5.90
Q6	1.322	1.05	1.04	0.47	-1.38	-0.68	2.34	5.01
Q8	0.229	0.78	0.78	0.67	-1.63	-0.94	0.62	2.86
Q9	-0.032	0.88	0.87	0.62	-2.32	-1.14	0.80	2.54
Q7	-0.059	1.05	1.05	0.50	-3.59	-0.83	0.87	3.31
Q3	-0.196	1.05	1.04	0.47	-3.64	-2.35	1.74	3.48
Q4	-0.606	0.87	0.86	0.58	-2.48	-1.76	-0.31	2.13
Q1	-0.636	0.93	0.87	0.57	-1.99	-1.00	-0.63	1.08
Q2	-0.653	0.89	0.84	0.58	-2.17	-1.00	-0.64	1.19
Q10	-0.971	0.85	0.82	0.61	-3.94	-1.37	-0.22	1.64

Table 3 shows that one item did not fit the Rasch PCM. Item 5 (Q5) did not fit the model because the Infit and Outfit MNSQs were outside the range of 0.5–1.5, and the PTMEA correlation was lower than 0.30. Infit and Outfit MNSQs indicated an underfit for item Q5. We believe that random or aberrant

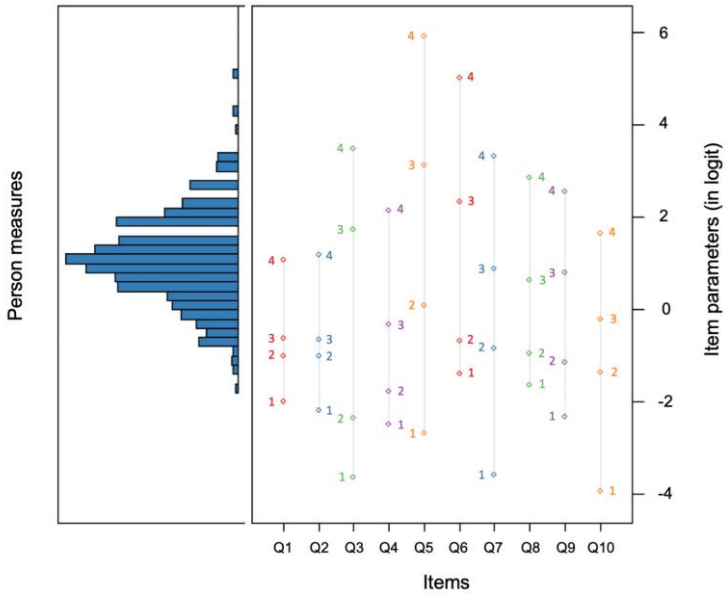
responses caused this, i.e. a person with a low level of self-control had the highest score for item 5. In addition, because of its low and negative PTMEA correlation, item Q5 was not sensitive enough to distinguish between a person with a low level of self-control and a person with a high level of self-control. Another possibility is that the translation from English to Indonesian is suboptimal and that this item is not very specific in measuring self-control in the Indonesian population. Furthermore, by inspecting the step parameter patterns, we found that step parameters for all items monotonically increased from low to high. This finding corroborates the monotonicity assumption of the model.

Wright Map

The Wright Map is one of the most significant innovations resulting from Rasch measurement. Using the Wright Map, the persons and items can be reviewed, and the relationship between persons and items can be inspected. Another aspect of the Wright Map is that persons and items are on the same scale, enabling insight into the respondents' performance on a set of test items (Liu & Boone, 2023). The Wright Map of the Indonesian BSCS is shown in Figure 2.

Figure 2

The Wright Map of the Indonesian BSCS



To complete the information in Figure 2, the mean of item location was 0.000, whereas the mean of person self-control level was 1.062. Thus, there was a 1-logit difference between the mean of the person measure and the mean of the item location. This finding indicates that in our study, person tended to have a higher level of self-control compared to the behavioural content of the items. Furthermore, the step parameters (1, 2, 3, and 4) of item Q5 spread widely across the continuum. However, item Q1 had a narrower step parameter range compared to Q5, Q3, Q6, and Q7. This finding is one of the reasons that PCM was found to be a better fit than the RSM since, in the latter, although the difference in the step parameters of Q5 and Q1 was very large, the threshold range was 'forced' to be mathematically uniform.

Reliability and separation indices

When using the Rasch model, reliability was reported for both item and person. Rasch-based reliability has two aspects: person separation reliability (PSR) and item separation reliability (ISR; Andrich & Marais, 2019). PSR is a measure of how well the measuring instrument differentiates between individuals with high ability and those with low ability and a measure of internal consistency, while ISR is a measure of how reliable the sample size is in classifying items in the hierarchy (Wright & Stone, 1999). A low ISR value indicates the need to increase the sample size so that the item hierarchy can be trusted. The PSR and ISR of the BSCS were 0.71 and 1.00, respectively. The PSR > 0.70 indicated that the BSCS has fairly good (acceptable) internal consistency and the ISR > 0.90 confirmed the item hierarchy (Linacre, 2021).

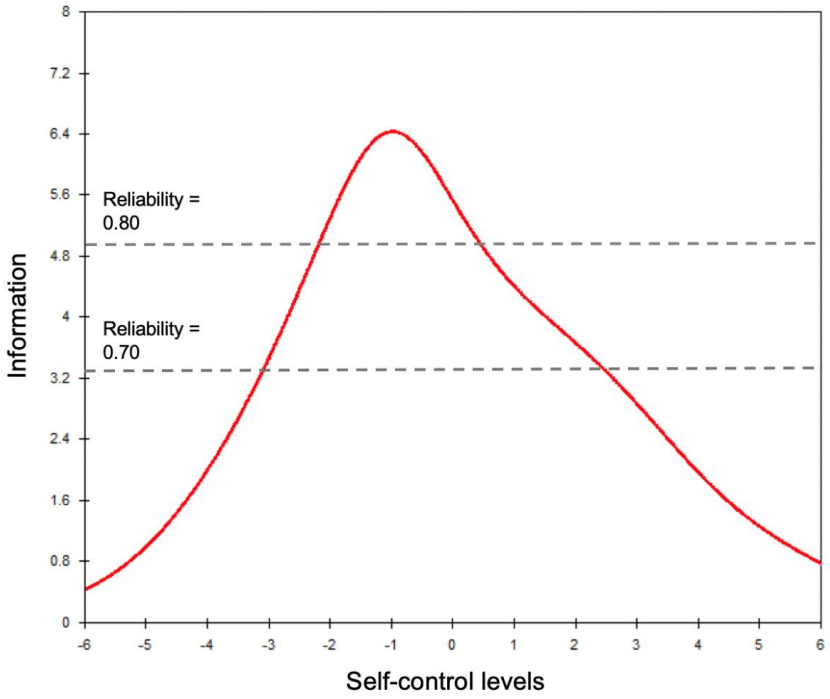
Test information function

The concept of the test information function (TIF) reflects how precisely a scale can measure the underlying trait, i.e. self-control (Wu et al., 2016). More specifically, TIF provides an explanation of the range of abilities for which the test provides the most precise measurement or produces the lowest measurement error when reliability for each ability level, called conditional reliability, can be calculated (DeMars, 2010). In the case of the PCM TIF of the Indonesian BSCS, the estimation results showed that for an

amount of information of 3.34, reliability was 0.70, while for 5.01, reliability was 0.80 (see Figure 3).

Figure 3

Test Information Function Curve of the Indonesian BSCS



Based on the estimation of the conditional reliability of the Indonesian BSCS, self-control was found to range from −3.061 to −2.152 and from 0.432 to 2.446 (71.82% of our study sample was in this range), resulting in a conditional reliability of 0.700–0.799. Furthermore, in the self-control range of −2.143 to 0.423 logits (21.47% of our study sample was within this range), conditional reliability ranged from 0.800 to 0.844. Based on these findings, we concluded that the Indonesian BSCS had optimal measurement precision that covered a fairly wide area (-3.061 to 2.446 logits), with 93.3% of this study's sample being in this range (only 6.7% were outside the range of optimal measurement precision).

DIF Analysis

Differential item functioning exists when different groups of test takers at the same ability level have significantly different chances of answering a test item because the test interacts with off-trait characteristics (Wu et al., 2016). To examine whether the items were of comparable difficulty for different gender groups, we examined the presence of DIF depending on gender. Rasch-based DIF analysis or the Rasch–Welch t-test was used (Smith, 1994). The criterion of at least a 0.400 logit difference (DIF contrast) with a p-value < 0.05 was used for detecting DIF (Linacre, 2013). The results of the DIF analysis (Table 4) indicated that item Q5 ('Fun and having fun sometimes prevent me from finishing my work') showed substantial DIF because the DIF contrast of −0.429 was greater than the predefined criterion. This finding indicates that males have a higher chance of obtaining the highest score (scored 5) on item Q5 compared to females, despite the two groups having the same level of self-control. We suspect that the content of item Q5 (i.e., the phrase 'having fun') had different meanings for males and females, leading to DIF of item Q5. Lastly, item Q9 has significant DIF ($p = 0.028 < 0.050$), but because the DIF contrast is relatively low at 0.200 (< 0.400), this item is categorized as having negligible DIF.

Table 4

Results of the DIF Analysis (Male – Female)

Item	DIF contrast	Joint S.E.	*t*	*df*	*p*
Q1	0.000	0.000	0.01	879	1.000
Q2	0.115	0.094	1.23	872	0.218
Q3	0.053	0.112	0.47	894	0.637
Q4	0.094	0.101	0.93	880	0.350
Q5	-0.425	0.112	-3.80	890	0.000
Q6	-0.029	0.102	-0.28	887	0.776
Q7	-0.021	0.096	-0.22	887	0.826
Q8	0.022	0.090	0.24	883	0.807
Q9	0.200	0.091	2.20	883	0.028
Q10	-0.110	0.094	-1.16	887	0.245

Discussion

The primary goal of this study was to assess the psychometric properties of the Indonesian BSCS using Rasch polytomous models, followed by DIF analysis to assess whether the Indonesian BSCS worked equally well across different gender groups. Based on our initial analysis, we compared the PCM and RSM and found that the PCM parameterisation fit better than the RSM. Such a comparison between the models is in line with previous studies (e.g. Gori et al., 2022; Youngerman et al., 2021). In addition, the Wright Map showed that the step parameter structure of the PCM was not uniform when there were large differences between item Q5 and item Q1. If the RSM were used, the differences in step parameters were forced to be equal. Therefore, the RSM had a poorer fit compared to the PCM. These findings are in line with Wu et al. (2016), who stated that the data rarely fit the RSM due to the assumption of an equal (uniform) step parameter structure for all items.

Based on unidimensionality or internal structure, we confirm that the Indonesian BSCS has a unidimensional factor structure. This factor structure is in line with the first adaptation study (Arifin & Milla, 2020), the original version of the scale (Tangney et al., 2004), and a recent study that also employed the Rasch model for the BSCS in China (Chen et al., 2022). However, this finding is not in line with the results of other studies that applied multidimensional IRT to the BSCS (Manapat et al., 2021) or the previous 10-item version of the BSCS (de Ridder et al., 2011). The 'disagreement' about the BSCS factor structure has been discussed in previous studies (e.g. Manapat et al., 2021).

The Indonesian BSCS did not show local dependence in the local independence test. This finding is in line with Chiesi et al.'s (2020) study that found that almost no BSCS models were modified by freeing the residual correlation in the CFA model, as well as studies that used modern test theory which did not find local dependence problems in the BSCS (Chen et al., 2022; Manapat et al., 2021). In addition, the Indonesian BSCS adaptation study did not modify the model by freeing residual correlation to achieve a model with a good fit (Arifin & Milla, 2020).

We found that all items had step parameters with a monotonically increasing pattern from lower to higher self-control levels. This finding indicates that the monotonicity assumption was met as recommended by previous studies (i.e., Tennant & Küçükdeveci, 2023). Therefore, the category functioning of the Indonesian BSCS does not experience rating scale malfunctioning (Wind, 2023) or disordered thresholds (Andrich, 2013), indicating that all response categories are well-functioning.

However, fit statistics showed that one item, Q5 ('*Hal yang menyenangkan dan bersenang-senang kadang menahan saya untuk menyelesaikan pekerjaan*' [Pleasure and fun sometimes keep me from getting work done]), did not fit the model. This item showed the greatest difficulty in obtaining the highest score. The reason for the poor fit of this item may be associated with random or aberrant responses (e.g., Karabatsos, 2000). We suspect that a number of respondents with a high level of self-control obtained the lowest score (a score of 1) on this item. Conversely, respondents with a low level of self-control may obtain the highest score (a score of 5) on this item. This condition causes fit statistics to experience strain (Outfit MNSQ > 1.5, PTMEA < 0.30) (Karabatsos, 2000). The procedure that can be used to obtain statistical evidence for this is response pattern analysis (Wright & Stone, 1999). In addition, we identified several flagged misfits, i.e. persons whose response patterns were suspected of causing the lack of fit of item Q5 (see Table S2 in Supplementary Materials). However, we did not conduct any follow-up analyses (i.e., analyses after removing misfits) because it was outside the focus of this study.

In the item hierarchy, the easiest item to agree with was Q10, '*Saya menolak hal-hal yang buruk untuk diri saya*' [I refuse things that are bad for me]. The review of the item content showed that the behaviour measured through this item was indeed very normative. However, because the application of the Rasch model to the BSCS is not widely studied, we cannot compare this level of difficulty with that of other studies. Finally, an ISR of 1.000 means that there is no problem with the item hierarchy or item spread (Wright & Stone, 1999).

Furthermore, the DIF analysis showed a DIF of Q5, indicating that this item tended to benefit males. This finding complements those of previous studies that focused on gender differences in self-control (Gibson et al.,

2010; Jo & Bouffard, 2014) but is inconsistent with the findings of other studies that the BSCS is invariant across gender (Chiesi et al., 2020). Although from a Rasch perspective, this method would help non-specialists to use the raw-score-to-logit conversion table (Saggino et al., 2020), when DIF is present, the raw score is no longer sufficient for the Rasch model (Linacre, 1992). Therefore, we did not report the raw-score-to-logit conversion table of the Indonesian BSCS. Further studies with other samples are needed to decide whether the DIF of Q5 item occurs consistently; if so, then dropping the Q5 item could be considered. The Indonesian BSCS had good internal consistency (PSR = 0.71). Our findings align with the results of other studies that applied the Rasch model to the BSCS (e.g. Chen et al., 2022). Based on the TIF, we found that the Indonesian BSCS covered a wide range of self-control levels classified as the optimal measurement precision range (high conditional reliability); 93.3% of our study sample had measures within the good measurement precision range. However, it should be noted that conditional reliability has a different meaning for PSR (single score) and should not be compared. Our finding also reflects the superiority of the Rasch model over CTT or CFA because of the model generates a conditional standard error of measurement for each level of trait levels (e.g. Andrich & Marais, 2019; Rusch et al., 2017).

This study has some limitations, both theoretical and methodological. The main theoretical limitation concerns the chosen BSCS model from a group of models - the 'Maloney model,' 'de Ridder model,' and 'Ferrari model' (e.g. Chen et al., 2022; Chiesi et al., 2020). While testing different models within a single study would be ideal, we only collected data using the 10-item version (i.e., the de Ridder version). Future studies should test various models using the 13-item version of the BSCS. The methodological limitation concerns administering the BSCS simultaneously with the pornography addiction test, which was the main focus of the broader research project. We believe that social desirability or response faking potentially had an impact on the Indonesian BSCS because respondents might have assumed that the BSCS instrument was related to the simultaneously administered pornography addiction instrument. Future studies should focus specifically on measuring self-control with the BSCS so that respondents are not distracted by other (negative) constructs.

Conclusion

In conclusion, this study is the first to validate and evaluate the psychometric properties of the Indonesian BSCS using the Rasch model. It showed that nine of the 10 Indonesian BSCS items were valid for measuring self-control. This study also showed that the basic assumptions of modern test theory - unidimensionality, local independence, and monotonicity - were fulfilled for this instrument. Lastly, our study should be replicated with different samples but applying the same method and same sample characteristics (i.e., high school students) to confirm whether DIF and misfit occur in item Q5.

Conflict of interest

We have no conflicts of interest to disclose.

Data availability statement

Data files are available upon a reasonable request.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561–573. https://doi.org/10.1007/BF02293814

Andrich, D. (2013). An expanded derivation of the threshold structure of the polytomous Rasch model that dispels any "threshold disorder controversy". *Educational and Psychological Measurement, 73*(1), 78–124. https://doi.org/10.1177/0013164412450877

Andrich, D., & Marais, I. (2019). *A course in rasch measurement theory: measuring in the educational, social and health sciences*. Springer Nature Singapore Pte Ltd. https://doi.org/10.1007/978-981-13-7496-8

Arifin, H., & Milla, M. (2020). Adaptasi dan properti psikometrik skala kontrol diri ringkas versi Indonesia. *Jurnal Psikologi Sosial, 18*(2), 179-195. https://doi.org/10.7454/jps.2020.18

Asgarian, F. S., Namdari, M., & Soori, H. (2020). Reliability and validity of Persian version of Brief Self-Control Scale (BSCS) in motorcyclists. *International Journal of Psychiatry in Clinical Practice, 24*(2), 176–182. https://doi.org/10.1080/13651501.2019.1711423

Bertrams, A., & Dickhäuser, O. (2009). High-school students' need for cognition, self-control capacity, and school achievement: Testing a mediation

hypothesis. *Learning and Individual Differences, 19*(1), 135–138. https://doi.org/10.1016/j.lindif.2008.06.005

Boone, W. J. (2020). Rasch basics for the novice. In M. S. Khine (Ed.), *Rasch measurement: Applications in quantitative educational research* (pp. 9-30). Springer. https://doi.org/10.1007/978-981-15-1800-3_2

Brevers, D., Foucart, J., Verbanck, P., & Turel, O. (2017). Examination of the validity and reliability of the French version of the Brief Self-Control Scale. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement, 49*(4), 243–250. https://doi.org/10.1037/cbs0000086

Chen, W., Zhang, G., Tian, X., & Zhao, S. (2022). Factor structure and longitudinal measurement invariance of the Tangney's Brief Self-Control Scale in Chinese adolescents. *Frontiers in Public Health, 10*, Article 802448. https://doi.org/10.3389/fpubh.2022.802448

Chiesi, F., Bonacchi, A., Lau, C., Tosti, A. E., Marra, F., & Saklofske, D. H. (2020). Measuring self-control across gender, age, language, and clinical status: A validation study of the Italian version of the Brief Self-Control Scale (BSCS). *PLoS ONE, 15*(8), e0237729. https://doi.org/10.1371/journal.pone.0237729

Cobb-Clark, D. A., Kong, N., & Schildberg-Hörisch, H. (2023). The stability of self-control in a population-representative study. *Journal of Economic Psychology, 95*, 1–10. https://doi.org/10.1016/j.joep.2022.102599

de Ridder, D. T. D., De Boer, B. J., Lugtig, P., Bakker, A. B., & van Hooft, E. A. J. (2011). Not doing bad things is not equivalent to doing the right thing: Distinguishing between inhibitory and initiatory self-control. *Personality and Individual Differences, 50*(7), 1006-1011. https://doi.org/10.1016/j.paid.2011.01.015

DeMars, C. (2010). *Item response theory*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195377033.001.0001

Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality, 45*(3), 259–268. https://doi.org/10.1016/j.jrp.2011.02.004

Fekih-Romdhane, F., Postigo, Á., González-Nuevo, C., Dabbous, M., Malaeb, D., Obeid, S., & Hallit, S. (2022). Psychometric properties of the Arabic version of the Oviedo Grit Scale (A-EGO) in non-clinical adults from the general population. *BMC Psychiatry, 22*(1), 792. https://doi.org/10.1186/s12888-022-04466-1

Ferrari, J. R., Stevens, E. B., & Jason, L. A. (2009). The role of self-regulation in abstinence maintenance: Effects of communal living on self-regulation. *Journal of Groups in Addiction & Recovery, 4*(1–2), 32–41. https://doi.org/10.1080/15560350802712371

Fung, S., Kong, C. Y. W., & Huang, Q. (2020). Evaluating the dimensionality and psychometric properties of the brief self-control scale amongst Chinese University Students. *Frontiers in Psychology*, *10*, 2903. https://doi.org/10.3389/fpsyg.2019.02903

García-Castro, F. J., Bendayan, R., & Blanca, M. J. (2024). Validity evidence for the Brief Self-Control Scale in the Spanish adult population: A systematic study. *Journal of Personality Assessment, 106*(2), 208-217. https://doi.org/10.1080/00223891.2023.2220403

Gibson, C. L., Ward, J. T., Wright, J. P., Beaver, K. M., & Delisi, M. (2010). Where does gender fit in the measurement of self-control? *Criminal Justice and Behavior, 37*(8), 883–903. https://doi.org/10.1177/0093854810369082

Gillebaart, M. (2018). The 'operational' definition of self-control. *Frontiers in Psychology, 9*, 1231. https://doi.org/10.3389/fpsyg.2018.01231

Gori, A., Topino, E., Svicher, A., Schuldberg, D., & Di Fabio, A. (2022). Insight orientation scale: A promising tool for organizational outcomes–A psychometric analysis using item response theory. *Frontiers in Psychology, 13*, 987931. https://doi.org/10.3389/fpsyg.2022.987931

Hagger, M., Zhang, C.-Q., Kangro, E.-M., Ries, F., Wang, J. C. K., Heritage, B., & Chan, D. K. C. (2021). Trait self-control and self-discipline: Structure, validity, and invariance across national groups. *Current Psychology, 40*(3), 1015-1030. https://doi.org/10.1007/s12144-018-0021-6

Holster, T. A., & Lake, J. (2016). Guessing and the Rasch model. *Language Assessment Quarterly, 13*(2), 124–141. https://doi.org/10.1080/15434303.2016.1160096

Irribarra, D. T., & Freund, R. (2024). Package 'WrightMap'. Retreived from https://cran.r-project.org/web/packages/WrightMap/WrightMap.pdf

Jo, Y., & Bouffard, L. (2014). Stability of self-control and gender. *Journal of Criminal Justice, 42*(4), 356–365. https://doi.org/10.1016/j.jcrimjus.2014.05.001

Karabatsos, G. (2000). A critique of Rasch residual fit statistics. *Journal of Applied Measurement, 1*(2), 152–176.

Kupper, N., Duijndam, S., & Karreman, A. (2020). Emotion expressivity in the Dutch: Validation of the Dutch translation of the Berkeley Expressivity

Questionnaire. *Psychological Assessment, 32*(5), 28–34. https://doi.org/10.1037/pas0000812

Linacre, J. (1992). Why fuss about statistical sufficiency? *Rasch Measurement Transactions, 6*(3), 230. https://www.rasch.org/rmt/rmt63c.htm

Linacre, J. M. (2013). Differential item functioning (DIF) sample size nomogram. *Rasch Measurement Transactions, 26*(4), 1391. https://www.rasch.org/rmt/rmt264a.htm

Linacre, J. M. (2021). *Winsteps® Rasch measurement computer program user's guide* (Version 5.1.4). Winsteps.com.

Liu, X., & Boone, W. J. (2023). Introduction to advances in applications of Rasch measurement in science education. In X. Liu, & W. J. Boone (Eds.), *Advances in applications of Rasch measurement in science education* (pp. 1-18). Springer. https://doi.org/10.1007/978-3-031-28776-3_1

Mair, P. (2018). *Modern psychometrics with R*. Springer International Publishing AG. https://doi.org/10.1007/978-3-319-93177-7

Maloney, P. W., Grawitch, M. J., & Barber, L. K. (2012). The multi-factor structure of the Brief Self-Control Scale: Discriminant validity of restraint and impulsivity. *Journal of Research in Personality, 46*(1), 111–115. https://doi.org/10.1016/j.jrp.2011.10.001

Manapat, P. D., Edwards, M. C., MacKinnon, D. P., Poldrack, R. A., & Marsch, L. A. (2021). A psychometric analysis of the Brief Self-Control Scale. *Assessment, 28*(2), 395-412. https://doi.org/10.1177/1073191119890021

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174. https://doi.org/10.1007/BF02296272

Milyavskaya, M., Berkman, E. T., & de Ridder, D. T. D. (2019). The many faces of self-control: Tacit assumptions and recommendations to deal with them. *Motivation Science, 5*(1), 79–85. https://doi.org/10.1037/mot0000108

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9*(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8

Nebioglu, M., Konuk, N., & Eroglu, Y. (2012). The investigation of validity and reliability of the Turkish version of the Brief Self-Control Scale. *Klinik Psikofarmakoloji Bülteni-Bulletin of Clinical Psychopharmacology, 22*(4), 340-351. https://doi.org/10.5455/bcp.20120911042732

Ozaki, Y., Goto, T., Kobayashi, M., & Kutsuzawa, G. (2016). Reliability and validity of the Japanese translation of Brief Self-Control Scale (BSCS-J). *Shinrigaku Kenkyu: The Japanese Journal of Psychology, 87*(2), 144-154. https://doi.org/10.4992/jjpsy.87.14222

Papanikolopoulos, P., Mastrotheodoros, S., Papanikolopoulou, S. Eleni Parlapani, E., & Kaprinis, S. G. (2022) The brief self-control scale: Dimensionality and psychometric properties in Greek young adults. *European Journal of Developmental Psychology, 19*(6), 925-937. https://doi.org/10.1080/17405629.2021.1964070

Paramitha, D. S., & Ariani, L. (2024). Hubungan kontrol diri dengan stress kerja pada bidan di rumah sakit Banjarmasin. *MANUJU: Malahayati Nursing Journal, 6*(6), 2415-2425. https://doi.org/10.33024/mnj.v6i6.11124

Pechorro, P., DeLisi, M., Gonçalves, R. A., Quintas, J., & Hugo Palma, V. (2021). The brief self-control scale and its refined version among incarcerated and community youths: Psychometrics and measurement invariance. *Deviant Behavior, 42*(3), 425–442. https://doi.org/10.1080/01639625.2019.1684942

Pilcher, J. J., Morris, D. M., & Erikson, D. N. (2023). Self-control measurement methodologies: An integrative approach. *Psychological Reports, 126*(3), 1108–1129. https://doi.org/10.1177/00332941211067969

Qonita, M., & Herdi, H. (2023). Asesmen kontrol diri peserta didik dan implikasinya terhadap program bimbingan dan konseling di SMA. *Jurnal Mahasiswa Bimbingan dan Konseling, 9*(2), 331-341. http://dx.doi.org/10.31602/jmbkan.v9i2.11257

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research.

Rusch, T., Lowry, P. B., Mair, P., & Treiblmaier, H. (2017). Breaking free from the limitations of classical test theory: Developing and measuring information systems scales using item response theory. *Information & Management, 54*(2), 189-203. https://doi.org/10.1016/j.im.2016.06.005

Saggino, A., Molinengo, G., Rogier, G., Garofalo, G., Loera, B., Tommasi, M., & Velotti, P. (2020). Improving the psychometric properties of the dissociative experiences scale (DES-II): A Rasch validation study. *BMC Psychiatry, 20*, 8. https://doi.org/10.1186/s12888-019-2417-8

Smith, E. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement, 3*(2), 205–231.

Smith, R. M. (1994). Detecting item bias in the Rasch Rating Scale Model. *Educational and Psychological Measurement, 54*(4), 886-896. https://doi.org/10.1177/0013164494054004004

Stavrova, O., Ren, D., & Pronk, T. (2022). Low self-control: A hidden cause of loneliness? *Personality and Social Psychology Bulletin, 48*(3), 347-362. https://doi.org/10.1177/01461672211007228

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality, 72*(2), 271–322. https://doi.org/10.1111/j.0022-3506.2004.00263.x

Tennant, A., & Küçükdeveci, A. A. (2023). Application of the Rasch measurement model in rehabilitation research and practice: early developments, current practice, and future challenges. *Frontiers in Rehabilitation Sciences, 4*, Article 1208670. https://doi.org/10.3389/fresc.2023.1208670

Welter, V. D. E., Dawborn-Gundlach, M., Großmann, L., & Krell, M. (2024). Adapting a self-efficacy scale to the task of teaching scientific reasoning: Collecting evidence for its psychometric quality using Rasch measurement. *Frontiers in Psychology, 15*, Article 1339615. https://doi.org/10.3389/fpsyg.2024.1339615

Wind, S. A. (2023). Detecting rating scale malfunctioning with the Partial Credit Model and Generalized Partial Credit Model. *Educational and Psychological Measurement, 83*(5), 953-983. https://doi.org/10.1177/00131644221116292

Wright, B. D., & Stone, M. (1999). *Measurement essentials* (2nd ed.). Wide Range, Inc.

Wu, M., Tam, H. P., & Jen, T.-H. (2016). *Educational measurement for applied researchers: theory into practice*. Springer Nature Singapore Pte Ltd.

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8*(2), 125-145. https://doi.org/10.1177/014662168400800201

Youngerman, E., Dahl, L. S., & Mayhew, M. J. (2021). Examining the psychometric properties of a New Integrative Learning Scale. *Research in Higher Education, 62*(6), 829-854. https://doi.org/10.1007/s11162-021-09623-1

Zwagery, R. V., Erlyani, N., Mayangsari, M. D., & Sugma, K. A. A. (2023). Kontrol diri dan perilaku pro-lingkungan. *Psibernetika, 16*(2), 63-69. http://dx.doi.org/10.30813/psibernetika.v16i2.3654

Supplementary Materials

Table S1

Descriptive Statistics for all Indonesian BSCS Items

Items	*M*	*SD*	Skewness	Minimum	Maximum
Q1	4.229	0.849	-1.103	1.000	5.000
Q2	4.206	0.841	-1.040	1.000	5.000
Q3	3.400	0.663	0.599	1.000	5.000
Q4	3.996	0.756	-0.535	1.000	5.000
Q5	2.820	0.666	0.017	1.000	5.000
Q6	3.033	0.740	-0.319	1.000	5.000
Q7	3.502	0.806	-0.177	1.000	5.000
Q8	3.594	0.876	-0.455	1.000	5.000
Q9	3.631	0.863	-0.213	1.000	5.000
Q10	4.061	0.817	-0.587	1.000	5.000

Figure S1

Plot of proportion of respondents for all response categories in all items

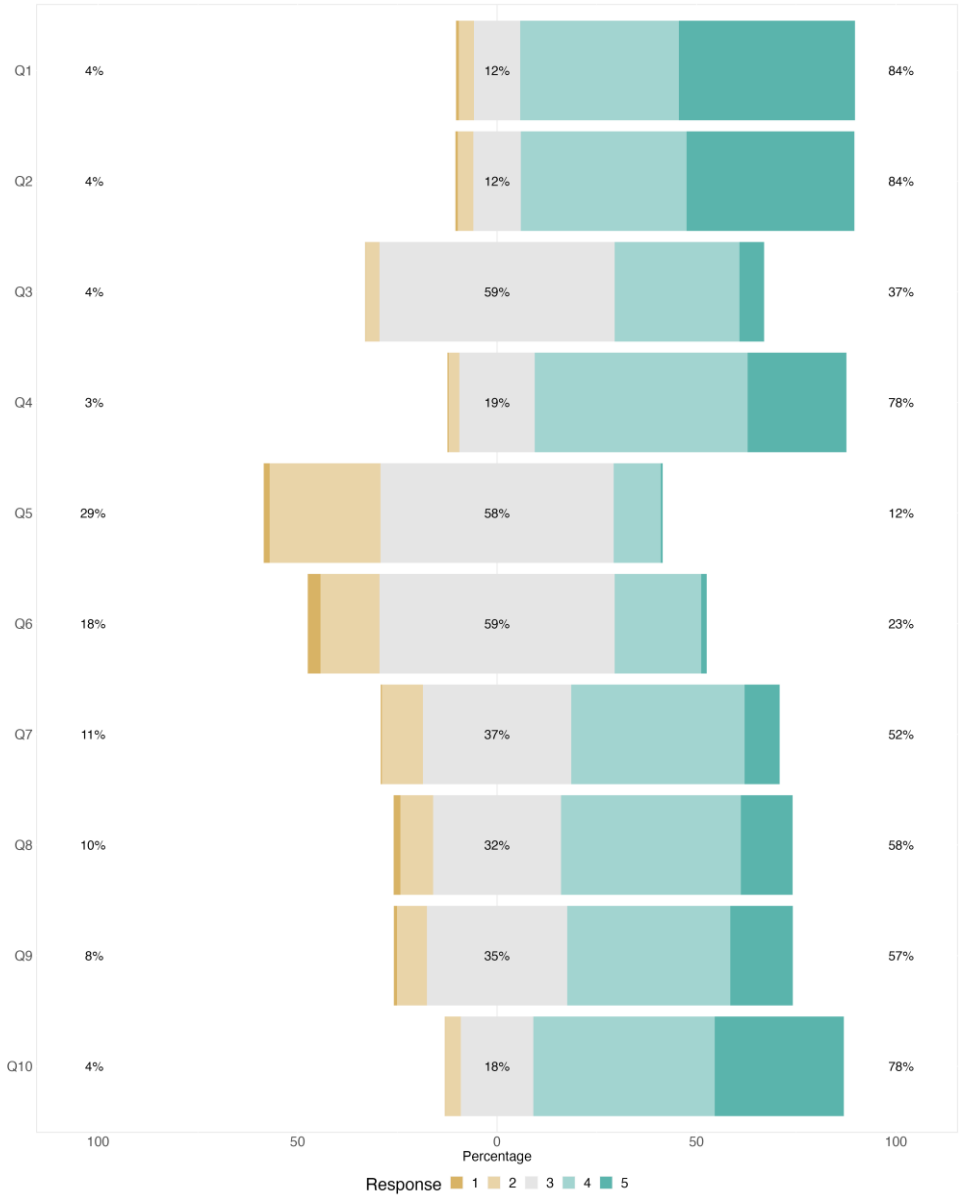


Table S2

Response pattern analysis for explaining misfit of Q5 (selected misfit persons)

Person ID	Theta	Response pattern										Outfit MNSQ	Outfit ZSTD
		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10		
P0654	2.385	5	5	3	5	1	5	4	5	4	5	3.22	4.1
P0293	1.828	5	3	4	4	1	5	4	4	5	5	2.88	2.7
P0832	2.099	5	5	3	5	1	4	5	4	4	5	3.05	2.9

Note. Table S1 shows three examples of our respondents who did not fit the partial credit model. These respondents had high levels of self-control, but instead responded with the lowest scores on Item Q5. This response pattern is what mathematically causes: (1) Outfit MNSQ of Q5 > 1.50 (underfit); (2) PTMEA low and negative.

Primenjena psihologija
Vol. 18, No. 2, pp. 255–279, 2025

Research Article

Unwarranted Trust and Unwarranted Punishment: Scientistic Beliefs Predict Support for Penalizing Science Skeptics

Petar Lukić [1]✉, Iris Žeželj [1]

[1] *LIRA Lab, Department of Psychology, Faculty of Philosophy, University of Belgrade, Serbia*

ABSTRACT

Contemporary society demands laypeople to articulate their attitudes towards science. Whilst distrust in science undeniably has detrimental outcomes, we argue that its opposite, uncritical trust in science and scientists (i.e., scientism) also leads to potential unwarranted societal polarization. In Study 1, we observed that people who endorsed scientistic beliefs supported discriminatory policies against people who ignore scientific knowledge or promote anti-scientific views, ranging from restrictions on their media appearance to imprisonment and denial of healthcare. Study 2 replicated the effect and tested its potential mechanism: the tendency to moralize rationality mediated the relationship between scientistic beliefs and support for penalizing measures. Our findings demonstrate that unwarranted beliefs, even if desirable in their moderate version, are associated with increased support for unwarranted measures against dissenters.

Keywords: scientism, trust in science, trust in scientists, science skepticism, penalization

UDK: 001.894
DOI: 10.19090/pp.v18i2.2592
Received: 14.11.2024.
Revised: 16.12.2024.
Accepted: 19.12.2024.

✉ Corresponding author's email: petar.lukic@f.bg.ac.rs

Introduction

"In science we trust."

Sticker/Yard sign for sale on Amazon, 2024

The COVID-19 pandemic showcased that debates between people who trust science (*the believers*) and those who ignore official recommendations or even support non-scientific accounts *(the skeptics*) can be quite heated. In November 2021, two groups of Australian citizens protested: one vehemently opposing COVID-19 vaccines and the other supporting compulsory vaccination with similar passion (Kelley, 2021). Twitter exchanges on COVID-19 between believers and skeptics typically contained offensive language (Liao et al., 2023). Despite inadequate expertise, laypeople strongly advocated for or against medical recommendations (e.g., wearing masks, vaccination), often accusing their opponents of ignorance or worse. Heated debates about scientific issues are, of course, not reserved for public health issues; Moernaut et al. (2022) documented the exchange between global warming skeptics and believers, in which they accused each other of lack of intelligence, irrationality, or immorality.

Distrusting science may hinder societal progress and endanger the safety of individuals, as well as society as a whole. For example, distrust in science was found to be related to lower compliance with COVID-19 health recommendations, such as wearing masks or vaccination (Hromatko et al., 2021; Plohl & Musil, 2021). Climate change skepticism is also related to a lower willingness to behave in a way that would mitigate the effects of climate change (Huber et al., 2022; Liu et al., 2022; Spence et al., 2011). In comparison, the consequences of unwarranted trust in science are not so clear. We argue that understanding this mechanism may help in preventing further societal divisions. In this paper, we examine how uncritical trust in science and scientists relates to penalizing practices against science skeptics. While we fully agree that laypeople should delegate their trust to science and scientists and follow their recommendations, in light of their lack of expertise, it is worth asking how they become so confident or even radicalized that they argue for harsh treatment of science skeptics or even deny them basic rights. It could be that both extremes of (dis)trust in science

that lead to intolerance have a certain form of misunderstanding of the nature of scientific enterprise at their core. This misunderstanding is closely related to the philosophical concept of *scientism*.

Scientistic Beliefs

Scientism refers to the attribution of undisputed epistemic and moral status to science, a belief that scientific knowledge has no boundaries and can answer all relevant societal questions, including moral or existential ones (Haack, 2012; Stenmark, 2018). Scientistic views can also incorporate an idealized view of scientists, i.e., seeing them as absolutely truthful, motivated solely by knowledge acquisition, and, in general, exceptionally virtuous. These two aspects of scientistic views are conceptually separated: we can imagine a person who believes science is an infallible and all-powerful method for discovering the truth while also acknowledging that scientists, as human beings, are susceptible to accidental or deliberate mistakes. Empirically, uncritical trust in science and scientists are correlated but separate factors, which suggests their potential independent predictive value for various constructs (Lukić & Žeželj, 2024).

Recently, research showed that scientistic beliefs are related to dogmatism (Lukić & Žeželj, 2024), viewed as "a relatively closed cognitive organization of beliefs and disbeliefs about reality, organized around a central set of beliefs about absolute authority which, in turn, provides a framework for patterns of intolerance and qualified tolerance toward others" (Rokeach, 1954, p. 195). Such rigid organization of reality is typically related to intolerance to the ones opposing it — we thus believe that unwarranted idealization of science and scientists may carry the risk of blaming the science skeptics for setbacks of scientifically driven policies.

Trust in Science and Penalization of Science Skeptics

Deviations from the group norms call for measures to regulate behavior, such as ridicule, punishment, or even ostracism (Jetten & Hornsey, 2014; Schachter, 1951). Moreover, our need to regulate socially deviant behavior is shown to be stronger when the behavior seems to negatively affect our well-being, including our safety or health (Brauer & Chekroun, 2005). The norm of trusting science was especially prominent in the case of

the COVID-19 pandemic (Bicchieri et al., 2021). Several recent papers explored how science believers treat science skeptics, typically regarding vaccination. For example, vaccinated people were more likely to endorse fines for those who did not get vaccinated themselves or did not vaccinate their children or even to advocate for taking away their childcare benefits (Blanchard-Rohner et al., 2021). Similarly, in certain contexts, vaccinated people supported restricting basic human rights to non-vaccinated ones - for example, restricting healthcare for non-vaccinated people in case they get infected (Kasper et al., 2022). Vaccinated people also felt a specific type of pleasure (*Schadenfreude*) when presented with a scenario of an anti-vax physician dying due to COVID-19 complications (Barlett & Meier, 2023), which is in line with other similar findings about decreased levels of compassion for those not vaccinated (Claudy et al., 2022; Hatchman et al., 2024).

In addition to these penalizing actions being hurtful per se, advocating for such harsh measures can prove inefficient or even backfire. That is, such a way of communication might not be optimal for winning the skeptics over to trust science as it may further alienate them and entrench their positions (Henkel et al., 2023; Prosser et al., 2020).

Present Research

Across two studies, we tested whether support for penalizing measures against science skeptics related to scientistic beliefs, i.e., uncritical trust in science and scientists. In Study 1, we explored the support for penalizing measures against people who neglect scientific recommendations and their relationship with scientistic beliefs. In Study 2, we introduced the potential mechanism underlying this relationship, assuming moralizing rationality as its mediator. The study design and data collection were approved by the Institutional Review Board of the Department of Psychology, University of Belgrade (Protocol no. ##2021-100).

Study 1

We expected that both Uncritical trust in science and Uncritical trust in scientists would positively correlate with Support for penalization (H1.1). We also expected that both Uncritical trust in science and Uncritical trust in scientists would independently predict Support for penalization (H1.1a).

Method

Open science practices

The study is part of the first author's PhD thesis, and its design, hypotheses, and analyses were preregistered (https://aspredicted.org/V37_JPQ). All data and supplemental materials are available at https://osf.io/a9g7x/.

Sample

From the initial database of 272 entries, 67 participants were excluded (17 entries were incomplete, 34 participants failed attention checks, and 16 completed the questionnaires too quickly). Thus, a final sample of 205 psychology and sociology students (171 females, 33 males, one undeclared, M_{AGE} = 20.90, SD_{AGE} = 3.45) was retained. The planned sample size allowed us to detect correlations of r = .20 (80% power and p = .05).

Variables and instruments

Scientific beliefs

To measure *scientistic beliefs* we used the 20-topic Scientistic Beliefs Questionnaire (Lukić & Žeželj, 2024) containing a 12-topic Uncritical trust in science subscale (α = .65) and 8-topic Uncritical trust in scientists subscale (α = .67). Each topic represents a five-option Thurstone-type scale ranked from extremely scientistic, through moderately scientistic, balanced view of science/scientists, moderately antiscientific, to extremely antiscientific. For example, one of the topics is "The Possibility of Reaching the Truth," where the extremely scientistic claim is "Science can reach the absolute truth about everything that exists", and the extremely antiscientific

claim is “Science can never truly reach any truth.” The respondents were instructed to choose the option that captures their opinion best. To capture scientistic beliefs only, we awarded participants two points for each extremely scientistic answer and one point for each moderately scientistic answer; all other answers were scored zero, meaning the mean score range for both subscales was 0 to 2. We opted for the Thurstone-type scale to ensure we have grounds to claim there is a mid-score reflecting a balanced view towards science, and to make a finer distinction between unwarranted trust, balanced view, and distrust in science and scientists. To make sure these were truly reflected in the options we offered participants, the scale had gone through a five-stage expert validation process that included philosophers of science and psychometricians (for details on the construction procedure, consult Lukić & Žeželj, 2024). The full questionnaire is available at https://osf.io/tznk5.

Support for penalization

To measure Support for penalization, we constructed a novel 20-item questionnaire with a 5-point Likert-type scale. Items (Table 1) were related to topics such as healthcare restrictions (e.g., “Companies should not give paid sick leave to those who deliberately did not get vaccinated and then got sick”), traditional media and social media visibility restrictions (e.g., “People who question scientific facts should be restricted from appearing in the media”), and fines or prison sentencing (e.g., ”Dissemination of information that is contrary to scientific evidence should be punishable by a prison sentence”). Based on face validity, before the items were administered to participants, the initial pool of 25 items was shortened to 20 to avoid content overlap. The scale's internal consistency was high (α = .93). To test the structure of the instrument, we conducted principal component analysis; the loadings on the first component ranged from .37 to .81 (for details, consult Supplemental files). Additionally, confirmatory factor analysis showed single factor solution had excellent fit indices (*CFI* = .975, *TLI* = .972, *NFI* = .966, *GFI* = .972) except for χ2 (χ2(170) = 631.15, $p < .001$) and *RMSEA* (RMSEA = .11 [95% CI .11 - .13]) which slightly exceeded values for acceptable fit with all items loadings significant on the level of $p < .001$ (for details, consult Supplemental files). This is why we proceeded to calculate a single score on the Support for penalization scale (fully available at https://osf.io/a9g7x).

Results

Distributions of responses showed a significant portion of respondents were supportive of penalizing measures towards science skeptics, ranging from 4.9% for the most radical ones aimed directly at people (denying healthcare) to 50.3% for the ones aimed at media that platform the skeptical views (additional taxing) (Table 1). The average support for penalization across measures was 2.53 (*SD* = 0.83). As for the scientistic views, their average endorsement on a scale ranging from 0 to 2 was 0.37 (*SD* = 0.24) for Uncritical trust in science and 0.38 (*SD* = 0.30) for Uncritical trust in scientists.

In line with H1.1, both Uncritical trust in science ($r = .46$, $p < .001$) and Uncritical trust in scientists ($r = .18$, $p = .012$) were positively related to Support for penalization. As in the previous studies (Lukić & Žeželj, 2024), Uncritical trust in science and Uncritical trust in scientists were positively related ($r = .30$, $p < .001$).

To test if Uncritical trust in scientists predicted Support for penalization independently of Uncritical trust in science, we set Support for penalization as the dependent variable and introduced Uncritical trust in science in the first block and Uncritical trust in scientists in the second. The model was significant ($F(2,202) = 27.6$), explaining 21% of the total variance. Contrary to our expectations (H1.1a), only Uncritical trust in science proved to be a significant predictor ($\beta = .45$, $p < .001$); Uncritical trust in scientists did not independently contribute to the prediction ($\beta = .04$, $p = .52$).

We further examined relations between scientistic beliefs and specific penalizing measures. All but one behavior (i.e., denying medical care to unvaccinated people) listed in the Support for penalization questionnaire positively correlated with Uncritical trust in science, while around half of the behaviors positively correlated with Uncritical trust in scientists (Table 1). Both Uncritical trust in science ($r = .43$) and Uncritical trust in scientists ($r = .28$) correlated the most with the endorsement of the immediate deletion of anti-scientific internet comments.

Table 1

The Distribution of Answers to Penalizing Behaviors and Their Correlations with Scientistic Beliefs - Study 1

	Answers							Correlation (*r*)	
Item	1 Extremely Antiscientific	2	3	4	5 Extremely Scientistic	*M*	*SD*	Uncritical trust in science	Uncritical trust in scientists
Newspapers, magazines, and books promoting anti-scientific views and pseudoscience should be additionally taxed.	15.1%	18.0%	16.6%	28.8%	21.5%	3.23	1.37	.26***	.15*
This society should be much less tolerant of people who spread anti-scientific views.	14.6%	15.1%	22.0%	32.7%	15.6%	3.20	1.29	.41***	.17**
It should be prohibited by law to organize protests against scientifically proven practices, such as protests against compulsory vaccination of children.	25.4%	15.6%	22.4%	22.0%	14.6%	2.85	1.40	.31***	.16*
People who do not vaccinate their children simply because they do not believe in the effectiveness and safety of vaccines should be seriously financially sanctioned.	25.4%	17.6%	19.5%	22.4%	15.1%	2.84	1.42	.22***	.03
Persons who bring disrepute to science should not go without some kind of sanction.	19.5%	23.4%	24.4%	24.4%	8.3%	2.79	1.25	.31***	.17**

People who treat their children with homeopathy or other pseudoscientific means should be severely sanctioned.	24.9%	23.4%	17.1%	23.9%	10.7%	2.72	1.35	.33***	.05
Comments on the Internet written by people with anti-scientific views should be deleted without hesitation.	23.4%	22.4%	26.8%	16.6%	10.7%	2.69	1.29	.43***	.28***
Parents who do not believe in the safety of vaccines should not be allowed to send their children to kindergarten.	30.2%	17.6%	22.4%	12.2%	17.6%	2.69	1.46	.35***	.01
It would be good if there were some system of sanctioning people who advocate anti-scientific views.	24.4%	23.4%	21.0%	22.9%	8.3%	2.67	1.29	.34***	.14*
Those who write comments with anti-scientific content on the Internet should somehow be restricted from accessing public platforms such as Facebook, Twitter, YouTube, or online newspapers.	26.3%	24.9%	17.1%	21.5%	10.2%	2.64	1.35	.34***	.20**
Persons who advocate anti-scientific views deserve the public condemnation of the entire society.	25.9%	20.0%	27.8%	20.0%	6.3%	2.61	1.24	.32***	.06
People who criticize science should be restricted from appearing in the media.	25.4%	27.3%	19.0%	21.5%	6.8%	2.57	1.27	.33***	.20**
People who criticize scientists should not be given space in the media.	28.3%	31.2%	21.5%	15.6%	3.4%	2.35	1.15	.39***	.18**

People who do not get vaccinated despite the doctor's recommendation should be fined.	32.2%	28.3%	19.5%	14.1%	5.9%	2.33	1.23	.25***	.07
Dissemination of information that is contrary to scientific evidence should be punishable by a fine.	36.1%	28.3%	21.0%	10.2%	4.4%	2.19	1.16	.37***	.12*
Dissemination of information that is contrary to scientific evidence should be punishable by a prison sentence.	36.1%	28.3%	21.0%	10.2%	4.4%	2.19	1.16	.30***	.09
People who question scientific facts should be restricted from appearing in the media.	36.6%	29.8%	18.5%	9.8%	5.4%	2.18	1.18	.38***	.15*
People who do not vaccinate their children just because they do not believe in the effectiveness and safety of vaccines should be punished with prison.	46.3%	25.4%	15.1%	9.8%	3.4%	1.99	1.15	.13*	-.01
Companies should not give paid sick leave to those who got sick and deliberately did not get vaccinated.	59.0%	17.6%	13.7%	8.3%	1.5%	1.76	1.07	.19**	.05
The state should not medically treat those people who got sick and who deliberately did not get vaccinated.	70.2%	16.6%	8.3%	4.4%	0.5%	1.48	0.86	.07	.06

Note. The items are ordered by the average endorsement from the highest to the lowest.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Discussion

As expected, individuals with stronger scientistic beliefs were more inclined to express support for penalization of science skeptics. This was consistently observed across different penalizing measures - from additional taxing the media that promote anti-science, through restricting the right to protest against science-based practices, banning skeptics from the media, or even fining or imprisoning them for promoting anti-science, to fining and imprisoning people for disregarding official medical practices. On the other hand, we only found the hypothesized contribution of uncritical trust in science, not in scientists, indicating the former to be more crucial for understanding the support for penalization.

We may expect people uncritically trusting science to be more likely to deem rationality (well represented by scientific thinking) morally virtuous and thus more likely to penalize those who think or behave anti-scientifically, which is often considered irrational. There is some experimental evidence for this assumption: Ståhl et al. (2016) found that individuals prone to moralizing rationality are more likely to punish a fictional religious doctor for advising prayer to a patient who later died.

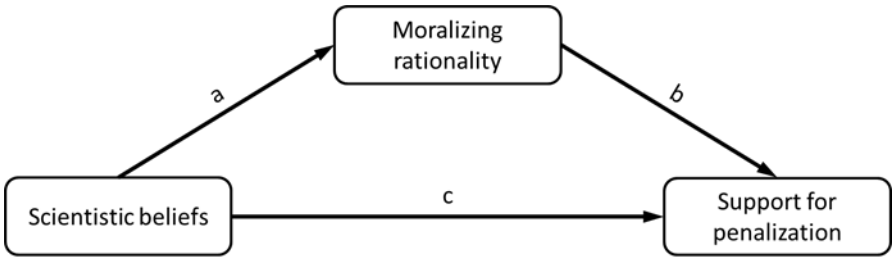
Study 2

In this study, we aimed to replicate the observed relationship between uncritical trust in science and support for discriminatory measures against skeptics in a more diverse sample. We also aimed to examine *moralizing rationality* as a potential underlying mechanism.

We expected that both Uncritical trust in science and Uncritical trust in scientists would correlate positively with the Support for penalization (H2.1). We also expected that moralizing rationality would mediate (Figure 1) the previously documented relationship between Uncritical trust in science and the Support for penalization (H2.2).

Figure 1

Mediation model



Method

Open science practices

The design of Study 2, hypotheses, and analyses were preregistered (https://aspredicted.org/3M5_HZW). All data and supplemental files are available at https://osf.io/a9g7x.

Sample

From the initial pool of 472 entries, after preregistered exclusions (231 entries were incomplete, 19 failed attention checks, and 5 completed the questionnaires too quickly), we were left with a total of N = 217 participants (142 females, 75 males, M_{AGE} = 39.13, SD_{AGE} = 11.79). They were recruited via posts on social media (i.e., Facebook and Instagram) and snowballing techniques. The sample was composed mostly of highly educated participants — 52.1% held an MA degree or higher, 27.2% had a university/college degree, 8.8% were students, and 13.0% had only a high school education.

Variable and instruments

To assess scientistic beliefs and the support for penalizing measures against skeptics, we used the same instruments as in Study 1. Uncritical trust

in science (α = .82), Uncritical trust in scientists (α = .77), and Support for penalization (α = .95) showed satisfactory internal consistencies.

We also used the 9-item Moralized rationality scale by Ståhl et al. (2016) with a 7-point (1- *Completely disagree*, 7 - *Completely agree*) Likert scale (e.g., "Being skeptical about claims that are not backed up by evidence is a moral virtue"), translated and adapted to Serbian via a parallel translation with consultations (α = .79).

Results

The frequencies of answers to Support for penalization items showed an even stronger endorsement of penalizing measures toward science skeptics than in Study 1. The average support for penalization was 3.09 (*SD* = 0.98). The average endorsement for Uncritical trust in science was 0.52 (*SD* = 0.37) and 0.43 (*SD* = 0.37) for Uncritical trust in scientists, both higher than in Study 1. Finally, people were moderately prone to moralize rationality (*M* = 4.02, *SD* = 1.13 on a scale from 1 to 7).

Table 2

The Distribution of Answers to Penalizing Behaviors and Their Correlations with Scientistic Beliefs - Study 2

	Answers							Correlations (*r*)	
Item	1 Extremely Antiscientific	2	3	4	5 Extremely Scientistic	*M*	*SD*	Uncritical trust in science	Uncritical trust in scientists
Newspapers, magazines, and books promoting anti-scientific views and pseudoscience should be additionally taxed.	13.4%	11.1%	12.4%	19.8%	43.3%	3.69	1.45	.45***	.36***
This society should be much less tolerant of people who spread anti-scientific views.	13.8%	9.7%	15.2%	25.8%	35.5%	3.59	1.41	.44***	.38***

Persons who bring disrepute to science should not go without some kind of sanction.	14.3%	9.7%	26.3%	20.7%	29.0%	3.41	1.37	.33***	.27***
Dissemination of information that is contrary to scientific evidence should be punishable by a fine.	13.4%	15.7%	19.8%	22.6%	28.6%	3.37	1.39	.45***	.41***
It would be good if there were some system of sanctioning people who advocate anti-scientific views.	16.1%	14.3%	19.4%	22.1%	28.1%	3.32	1.43	.39***	.37***
Persons who advocate anti-scientific views deserve the public condemnation of the entire society.	16.1%	15.7%	18.4%	23.0%	26.7%	3.29	1.42	.42***	.37***
Parents who do not believe in the safety of vaccines should not be allowed to send their children to kindergarten.	19.8%	14.3%	19.4%	18.4%	28.1%	3.21	1.49	.34***	.26***
People who do not vaccinate their children simply because they do not believe in the effectiveness and safety of vaccines should be seriously financially sanctioned.	21.7%	10.6%	22.1%	18.9%	26.7%	3.18	1.49	.31***	.34***
It should be prohibited by law to organize protests against scientifically proven practices, such as protests against compulsory vaccination of children.	21.2%	14.8%	18.0%	18.0%	28.1%	3.17	1.51	.37***	.30***

People who criticize science should be restricted from appearing in the media.	17.5%	17.5%	19.4%	26.3%	19.4%	3.12	1.38	.37***	.26***
Those who write comments with anti-scientific content on the Internet should somehow be restricted from accessing public platforms such as Facebook, Twitter, YouTube, or online newspapers.	20.7%	17.5%	21.2%	18.9%	21.7%	3.03	1.44	.35***	.36***
Comments on the Internet written by people with anti-scientific views should be deleted without hesitation.	20.7%	18.0%	24.0%	16.6%	20.7%	2.99	1.42	.36***	.36***
People who treat their children with homeopathy or other pseudoscientific means should be severely sanctioned.	23.5%	18.4%	17.5%	18.0%	22.6%	2.98	1.49	.37***	.25***
People who do not get vaccinated despite the doctor's recommendation should be fined.	29.0%	13.8%	20.7%	19.8%	16.6%	2.81	1.46	.27***	.30***
Dissemination of information that is contrary to scientific evidence should be punishable by a prison sentence.	25.4%	18.4%	24.4%	15.2%	16.6%	2.79	1.41	.43***	.33***
People who criticize scientists should not be given space in the media.	24.4%	18.9%	24.9%	18.4%	13.4%	2.77	1.36	.34***	.28***

People who question scientific facts should be restricted from appearing in the media.	28.6%	21.2%	23.0%	13.8%	13.4%	2.62	1.38	.34***	.30***
People who do not vaccinate their children just because they do not believe in the effectiveness and safety of vaccines should be punished with prison.	39.6%	17.5%	21.2%	10.6%	11.1%	2.36	1.38	.25***	.19**
Companies should not give paid sick leave to those who got sick and deliberately did not get vaccinated.	43.8%	16.6%	18.0%	10.6%	11.1%	2.29	1.40	.33***	.11
The state should not medically treat those people who got sick and who deliberately did not get vaccinated.	56.2%	17.1%	14.8%	4.6%	7.4%	1.90	1.25	.20**	.13*

Note. The items are ordered by the average endorsement from the highest to the lowest.

* $p < .05$, ** $p < .01$, *** $p < .001$.

As expected, the correlation between Uncritical trust in science and Uncritical trust in scientists was high ($r = .41$, $p < .001$). Support for penalization correlated positively with both Uncritical trust in science ($r = .49$, $p < .001$) and Uncritical trust in scientists ($r = .41$, $p < .001$), in line with H2.1. Moralized rationality correlated positively with all three other variables, more strongly with Support for penalization ($r = .43$, $p < .001$) and Uncritical trust in science ($r = .42$, $p < .001$) than with Uncritical trust in scientists ($r = .27$, $p < .001$).

An exploratory regression analysis (specified in the same way as in Study 1) now showed both Uncritical trust in science ($\beta = .39$, $p < .001$) and Uncritical trust in scientists ($\beta = .25$, $p < .001$) to independently predict Support for penalization ($F(2,214) = 44.6$), explaining 29% of its total variance.

Mediation analysis

The mediation analysis was conducted in the R *lavaan* package (R Core Team, 2021; Rosseel, 2012). The analysis (ML as an estimator) was conducted with 5,000 bootstrapped samples to estimate the indirect effect and its confidence intervals. The results revealed a significant positive effect of Uncritical trust in science on Support for penalization (c = .49, SE = .06, t = 8.34, p < .001). When Moralized rationality was introduced in the model, the direct effect of Uncritical trust in science on Support for penalization remained significant but was reduced (c' = .38, SE = .06, t = 6.82, p < .001). As expected (H2.2), the indirect path from Uncritical trust in science to Support for penalization through Moralized rationality was significant (ab = .11, SE = .03, t = 3.46, p < .001), with a 95% bootstrapped confidence interval [0.05, 0.18], indicating partial mediation. The model explained 30% of the variance of Support for penalization. For the complete set of regression coefficients, consult the Supplemental files.

While we did not preregister our hypothesis regarding the effect of Uncritical trust in scientists on Support for penalization through Moralized rationality, the effect was significant (ab = .09, SE = .03, t = 3.45, p < .001), with a 95% bootstrapped confidence interval [0.05, 0.15]. The direct effect was significant before (c = .41, SE = .06, t = 6.62, p < .001) and after the inclusion of Moralized rationality (c' = .32, SE = .05, t = 6.13, p < .001). This model explained 28% of the variance of Support for penalization.

Discussion

Study 2 corroborated the relationships between scientistic beliefs and support for penalizing measures against science skeptics. The correlations were even stronger than in Study 1, especially the one between Uncritical trust in scientists and Support for penalization. Moreover, the endorsement of penalizing measures was higher than in the previous study, which may be attributed to the increased diversity of the sample in terms of gender, age, and education. Participants agreed the most with the item suggesting that media promoting anti-scientific views and pseudoscience should be additionally taxed (63%). As in Study 1, participants agreed the least with denying medical care to people who deliberately did not get vaccinated but got ill (12.0%). On this micro level, we also observed stronger

correlations than in Study 1, as Uncritical trust in science correlated positively with all specific penalizing measures, and Uncritical trust in scientists with all but one (In Study 1, it only correlated with around half of these measures).

We further revealed individuals who uncritically trusted science and scientists more were also more likely to moralize rationality (i.e., to perceive scientific thought and rationality as a moral imperative). Thus, trust in science and scientists seems to have transcended from the domain of evidence and rationality to the domain of values, so violating these moral values further calls for sanctions. The proposed mediation effect, however, was only partial, leaving a significant direct path from scientistic beliefs to penalizing measures. Since no causal relations could be inferred, this suggests that broader beliefs about the moral value of being rational could explain the relationship between scientistic beliefs and the support for penalization of science skeptics, but that there is still a remaining shared variance specific to the relationship between beliefs about science and penalization.

Our results suggest that uncritical trust in science plays a more central role in driving support for the penalization of science skeptics than uncritical trust in scientists. This could be partially due to the fact that the former correlates with moralizing rationality more strongly than the latter. It suggests that scientific thought is held as more "sacred" than its actual practitioners, i.e., violations of scientific thought are more likely to elicit penalization than distrust in scientists themselves. We have also observed an incremental contribution of Uncritical trust in scientists to the prediction of penalization only in Study 2 and not in Study 1. One possible explanation is that the distribution of this measure was better in the community sample, compared to the student sample in Study 1, in which its range was more restricted.

General Discussion

Across two studies, we compellingly showed that people who deemed science and scientists supreme were also more prone to endorse different penalizing measures against people who ignore scientific recommendations or advocate for anti-scientific views.

We found that a significant portion of respondents in both studies supported each of the penalization items. Even very harsh sanctions, such as imprisonment or denying state-funded medical treatment, were endorsed by more than 10% of the respondents in the online community sample in Study 2. Around 46% of these respondents, the majority of whom are highly educated, agreed that skeptics should be prohibited from protesting, while around one-third thought that people who question scientific facts or criticize scientists should be prohibited from appearing in the media.

This nuanced look at the type of behaviors being penalized and the harshness of penalizing measures also revealed that scientistic beliefs related primarily to lighter forms of penalization, such as restricting their appearances in the media and heavy moderation of unscientific comments in digital space. Scientistic beliefs appeared to be less strongly related to more radical measures with health-related consequences, such as denying medical care to unvaccinated people and prison sentences to parents of unvaccinated children.

Our results, thus, show that extreme views of science are related to extreme views about how citizens should be legally regulated in the context of science-related societal issues. While it may seem reasonable to call for sanctions against some of these serious anti-scientific behaviors as they may seriously endanger our own and the well-being of others, the matter of considering and imposing sanctions would be a complex task for policymakers. For example, before introducing any sanctions against those spreading anti-scientific information, the notions of the right to free speech and the knowledge of the speaker (i.e., the differentiation between ignorance and false advertising, fraud, or hoax) should be considered (Gielow Jacobs, 2022). Also, due to civil rights, criminal law measures should only be used as the last option - when all other measures prove ineffective (Husak, 2004). Finally, legal sanctions do not necessarily prove efficient, as they often undermine the intentions to curb the consequences of anti-scientific behavior (Sun et al., 2022).

Adherence to science-based recommendations was previously shown to be moralized (e.g., Bor et al., 2023; Rozin & Singh, 1999; Salomon et al., 2017). This implies that non-adherence would consequently call for appropriate sanction. We thus hypothesized moralizing rationality might play

a role in the relationship between scientistic beliefs and the support for penalization of science skeptics. Moralizing rationality indeed partially explained the relationship between scientistic beliefs and the support for penalizing measures, suggesting a possible mechanism behind this relationship. However, other concepts could also be at the root of both scientistic beliefs and proneness to punish science skeptics. Clearly, this issue is not politics- or ideology-free, so political orientation probably plays a significant role in these relationships. The likely trait-like candidates could be dogmatism (Rokeach, 1954) and social dominance orientation — support for societal hierarchy (Pratto et al., 1994). For example, dogmatism could bridge scientistic beliefs and support for penalization by providing a wider mental framework characterized by intolerance against those who disobey the authority of science. Likewise, individuals who support social hierarchies may be more inclined to accept policies that limit the power and privileges of certain groups, especially if they view those groups' positions as unwarranted or morally wrong. However, an abundance of research shows science skepticism is stronger for those with right-leaning political attitudes (Gauchat, 2012; Hamilton et al., 2015; Mann & Schleifer, 2020). To prevent potential ideological confounds, one would thus need to use as ideology-free measures as possible, which is why we suggested trait-like ones, such as dogmatism. Finally, as these psychological processes do not happen in a vacuum, the associations we demonstrated could be even exacerbated if the context is perceived as threatening (as, for example, during the recent COVID-19 pandemic).

Limitations and Further Research

As this was the first look into the relationship between scientistic beliefs and penalization, we collected data on convenience samples. To assess the prevalence of endorsement of penalizing measures, a representative sample is needed. Next, as both the PCA and CFA were conducted on the same sample, future research should seek to validate the latent structure using independent samples to ensure the robustness and generalizability of the findings. The robustness of the observed correlations, however, is very promising, and we expect them to be replicated independently. Next, due to the correlational design, we cannot speak to the

direction of the effect. Future experimental studies should thus target scientistic beliefs and look at their subsequent effects. Other potential paths that lead to support for penalization could also be tested, for example, those originating from personality traits; the role of contextual moderators, such as a feeling of threat, might also help us better understand the observed relationships. Finally, we opted to measure policy endorsement; however, penalizing behaviors could also manifest themselves on an interpersonal level and may be measured either through self-reported (e.g., social distance towards science skeptics) or observed behavior (e.g., readiness to help, physical distancing, etc.).

Conclusion

While one cannot dispute that science skepticism endangers public health and erodes global efforts to transition to a more sustainable lifestyle, we argue that its opposite pole — idealizing science and being unaware of its limits — could backfire so that it deepens societal divisions and further alienates skeptics, making them even harder to reach. This finding adds to the existing literature showing that, even when socially desirable in its moderate form, any belief that becomes dogmatic and extreme may lead to detrimental societal consequences.

Acknowledgements

We thank Predrag Teovanović and Marija Petrović for their kind reading of the first draft of the manuscript.

Funding

The realization of this research was financially supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia as part of the financing of scientific research work at the University of Belgrade - Faculty of Philosophy (contract number 451-03-47/2023-01/ 200163).

Conflict of interest

We have no conflicts of interest to disclose.

Data availability statement

Data used in this paper can be found at https://osf.io/a9g7x.

References

Barlett, C. P., & Meier, B. P. (2023). Schadenfreude in the context of opposing vaccination statuses. *European Journal of Social Psychology, 53*, 1593–1604. https://doi.org/10.1002/ejsp.2998

Bicchieri, C., Fatas, E., Aldama, A., Casas, A., Deshpande, I., Lauro, M., ... & Wen, R. (2021). In science we (should) trust: Expectations and compliance across nine countries during the COVID-19 pandemic. *PloS One, 16*(6), e0252892. https://doi.org/10.1371/journal.pone.0252892

Blanchard-Rohner, G., Caprettini, B., Rohner, D., & Voth, H. J. (2021). Impact of COVID-19 and intensive care unit capacity on vaccination support: Evidence from a two-leg representative survey in the United Kingdom. *Journal of Virus Eradication, 7*(2), 100044. https://doi.org/10.1016/j.jve.2021.100044

Bor, A., Jørgensen, F., Lindholt, M. F., & Petersen, M. B. (2023). Moralizing the COVID-19 pandemic: Self-interest predicts moral condemnation of other's compliance, distancing, and vaccination. *Political Psychology, 44*(2), 257-279. https://doi.org/10.1111/pops.12835

Brauer, M., & Chekroun, P. (2005). The relationship between perceived violation of social norms and social control: Situational factors influencing the reaction to deviance. *Journal of Applied Social Psychology, 35*(7), 1519-1539. https://doi.org/10.1111/j.1559-1816.2005.tb02182.x

Claudy, M. C., Vijayakumar, S., & Campbell, N. (2022). Reckless spreader or blameless victim? How vaccination status affects responses to COVID-19 patients. *Social Science & Medicine, 305*, 115089. https://doi.org/10.1016/j.socscimed.2022.115089

Gauchat, G. (2012). Politicization of science in the public sphere: A study of public trust in the United States, 1974 to 2010. *American Sociological Review, 77*(2), 167-187. https://doi.org/10.1177/0003122412438225

Gielow Jacobs, L. (2022). Freedom of speech and regulation of fake news. *The American Journal of Comparative Law, 70*(Supplement_1), i278-i311. https://doi.org/10.1093/ajcl/avac010

Haack, S. (2012). Six signs of scientism. *Logos & Episteme, 3*(1), 75-95. https://doi.org/10.5840/logos-episteme20123151

Hamilton, L. C., Hartter, J., & Saito, K. (2015). Trust in scientists on climate change and vaccines. *Sage Open, 5*(3), 2158244015602752. https://doi.org/10.1177/2158244015602752

Hatchman, K., Hornsey, M. J., & Barlow, F. K. (2024). The vaccination divide: Exploring moral reasoning associated with intergroup antipathy between vaccinated and unvaccinated people. *British Journal of Health Psychology, 29*(4), 889-906. https://doi.org/10.1111/bjhp.12736

Henkel, L., Sprengholz, P., Korn, L., Betsch, C., & Böhm, R. (2023). The association between vaccination status identification and societal polarization. *Nature Human Behaviour*, *7*(2), 231-239. https://doi.org/10.1038/s41562-022-01469-6

Hromatko, I., Tonković, M., & Vranić, A. (2021). Trust in science, perceived vulnerability to disease, and adherence to pharmacological and non-pharmacological COVID-19 recommendations. *Frontiers in Psychology*, *12*, 664554. https://doi.org/10.3389/fpsyg.2021.664554

Huber, R. A., Greussing, E., & Eberl, J. M. (2022). From populism to climate scepticism: the role of institutional trust and attitudes towards science. *Environmental Politics, 31*(7), 1115-1138. https://doi.org/10.1080/09644016.2021.1978200

Husak, D. (2004). The criminal law as last resort. *Oxford Journal of Legal Studies*, *24*(2), 207-235. https://doi.org/10.1093/ojls/24.2.207

Jetten, J., & Hornsey, M. J. (2014). Deviance and dissent in groups. *Annual Review of Psychology, 65*(1), 461-485. https://doi.org/10.1146/annurev-psych-010213-115151

Kasper, A., Frébert, N., & Testé, B. (2022). Caught COVID-19? Covidiot! Attributions of humanness and social punishment toward an ingroup member who breaks lockdown rules. *Social Psychology, 53*(2), 84–95. https://doi.org/10.1027/1864-9335/a000478

Kelley, L. (2021, November 20). Pro and anti-vaccination protesters take to Australia streets. *Reuters*. Retrieved from https://www.reuters.com/markets/asia/pro-anti-vaccination-protesters-take-australia-streets-2021-11-20/

Liao, S., Okpala, E., Cheng, L., Li, M., Vishwamitra, N., Hu, H., ... & Costello, M. (2023, August). *Analysis of COVID-19 offensive tweets and their targets*. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 4473-4484). https://doi.org/10.1145/3580305.3599773

Liu, T., Shryane, N., & Elliot, M. (2022). Attitudes to climate change risk: classification of and transitions in the UK population between 2012 and

2020. *Humanities and Social Sciences Communications, 9*(1), 1-15. https://doi.org/10.1057/s41599-022-01287-1

Lukić, P., & Žeželj, I. (2024). Delineating between scientism and science enthusiasm: Challenges in measuring scientism and the development of novel scale. *Public Understanding of Science*, *33*(5), 568-586. https://doi.org/10.1177/09636625231217900

Mann, M., & Schleifer, C. (2020). Love the science, hate the scientists: Conservative identity protects belief in science and undermines trust in scientists. *Social Forces*, *99*(1), 305-332. https://doi.org/10.1093/sf/soz156

Moernaut, R., Mast, J., Temmerman, M., & Broersma, M. (2022). Hot weather, hot topic. Polarization and sceptical framing in the climate debate on Twitter. *Information, Communication & Society, 25*(8), 1047-1066. https://doi.org/10.1080/1369118X.2020.1834600

Plohl, N., & Musil, B. (2021). Modeling compliance with COVID-19 prevention guidelines: The critical role of trust in science. *Psychology, Health & Medicine*, *26*(1), 1-12. https://doi.org/10.1080/13548506.2020.1772988

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, *67*(4), 741–763. https://doi.org/10.1037/0022-3514.67.4.741

Prosser, A. M., Judge, M., Bolderdijk, J. W., Blackwood, L., & Kurz, T. (2020). 'Distancers' and 'non-distancers'? The potential social psychological impact of moralizing COVID-19 mitigating practices on sustained behaviour change. *British Journal of Social Psychology*, *59*(3), 653-662. https://doi.org/10.1111/bjso.12399

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rokeach, M. (1954). The nature and meaning of dogmatism. *Psychological Review*, *61*(3), 194-204. https://doi.org/10.1037/h0060752

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1-36. https://doi.org/10.18637/jss.v048.i02

Rozin, P., & Singh, L. (1999). The moralization of cigarette smoking in the United States. *Journal of Consumer Psychology*, *8*(3), 321-337. https://doi.org/10.1207/s15327663jcp0803_07

Salomon, E., Preston, J. L., & Tannenbaum, M. B. (2017). Climate change helplessness and the (de) moralization of individual energy behavior. *Journal of Experimental Psychology: Applied*, *23*(1), 15. https://doi.org/10.1037/xap0000105

Schachter, S. (1951). Deviation, rejection, and communication. *The Journal of Abnormal and Social Psychology, 46*(2), 190–207. https://doi.org/10.1037/h0062326

Spence, A., Poortinga, W., Butler, C., & Pidgeon, N. F. (2011). Perceptions of climate change and willingness to save energy related to flood experience. *Nature Climate Change, 1*(1), 46-49. https://doi.org/10.1038/nclimate1059

Ståhl, T., Zaal, M. P., & Skitka, L. J. (2016). Moralized rationality: Relying on logic and evidence in the formation and evaluation of belief can be seen as a moral issue. *PloS One*, 11(11), e0166332. https://doi.org/10.1371/journal.pone.0166332

Stenmark, M. (2018). Scientism and its rivals. In: J. De Ridder, R. Peels and R. Van Woudenberg (Eds.), *Scientism: Prospects and problems* (pp. 57–82). Oxford, Oxford University Press. https://doi.org/10.1093/oso/9780190462758.001.0001

Sun, N., Christie, E., Cabal, L., & Amon, J. J. (2022). Human rights in pandemics: criminal and punitive approaches to COVID-19. *BMJ Global Health*, *7*(2), e008232. https://doi.org/10.1136/bmjgh-2021-008232

Primenjena psihologija
Vol. 18, No. 2, pp. 281–309, 2025

Research Article

The Differences in the Importance of Upbringing Tolerance in a Chinese Family Across Generations

Liubov Ryumshina[1], Anastasiya Chernova[1], Elena Zinchenko[1], Ziyoda Abidova[2], Yuliya Berdyanskaya [2] ✉, and Olga Batychko[1]

[1] *Southern Federal University, Rostov-on-Don, Russian Federation*
[2] *National University of Uzbekistan named after Mirzo Ulugbek, Tashkent, Uzbekistan*

ABSTRACT

The article uses China as an example to demonstrate the importance and possibility of taking big data into account when studying intergenerational relationships during a period of great change. The aims of the study were to examine the importance of nurturing tolerance in children in Chinese families. To ensure representativeness, different generations of adults who live in diverse socio-economic conditions were included in the study. The study used the World Values Survey (WVS) database (China's participation from 1990 to 2018), and the sample included 14,489 respondents. According to the results of the frequency analysis, tolerance ranks 2-5 (depending on the year of the study) among the qualities important for nurturing in a Chinese family. The qualities that Chinese respondents rate as more important for raising children than tolerance include independence and diligence. The highest value for the importance of nurturing tolerance in the family was recorded in 2001, but since 2007, a statistically significant downward trend has been established. Respondents under 29 often rate tolerance higher than respondents aged 30-49 and over 50.

Keywords: tolerance, collectivistic cultures, family upbringing, generations

UDK: 159.924.7:316.647.5(510)
DOI: 10.19090/pp.v18i2.2562
Received: 27.06.2024.
Revised: 26.12.2024.
Accepted: 05.01.2025.

✉ Corresponding author's email: yber@sfedu.ru

Introduction

Tolerance is one of the most relevant and complex concepts of our time and is studied from the standpoint of various sciences. Tolerance is a consequence of living in conditions of diversity, which is an essential attribute of human existence. However, tolerance does not involve appreciating this diversity but rather recognizing the right of "others" (e.g., cultures, countries, and people who differ in religion, values, opinions, and behavior) to exist in the world. It is the right of other people to have different beliefs and customs as long as they do not violate common moral values, which may vary across cultures and societies. A significant number of works has been devoted to the connection between morality and tolerance (Armstrong & Wronski, 2019; Song, 2018), as well as the dependence of the latter on the norms of human existence (universal values; Hamer et al., 2019; Verkuyten & Killen, 2021).

Currently, interethnic tolerance is being studied the most (Park et al., 2022). Such an interest in this type of tolerance is due to globalization, which caused migration in the modern world, as well as the aggravation of the issue of terrorism and the increase in the number of refugees – also relevant at present. However, tolerance as a social phenomenon is an integral part of any human communication. At all its levels (intergroup and interpersonal), it affects how communication proceeds and how it ends: with understanding and acceptance or conflict.

In a social context, tolerance is a person's willingness to allow other people to choose their own lifestyle and behavior in the absence of negative (aggressive and violent) manifestations. Social tolerance is a non-violent, respectful relationship between different social groups (i.e., groups of people of different ages, financial situations, social statuses, subcultures, etc.). Tolerance in interpersonal relationships is a conscious, tolerant attitude toward a communication partner, which implies the recognition and respect of their right to manifest individuality. Thus, the manifestation of a tolerant attitude towards another person contributes to a free and open dialogue, often leading to agreement.

As it becomes clear, tolerance is a value necessary for constructive interaction in the human realm. Therefore, it is no coincidence that scientists

in all countries pay so much attention to its formation. The literature review showed that the development of tolerance was considered in numerous studies. Most of these studies examined the development of tolerance in public institutions (Miklikowska, Bohman & Titzmann, 2019): factors and conditions that help schools and teachers promote mutual understanding (Rjumshina, 2000; Ryumshina, Berdyanskaya et al., 2022; Shestakova et al., 2022) and tolerance in a globalized world (Bayram Özdemir & Özdemir, 2020; Brenik et al., 2019; Sandoval-Hernandez et al., 2018; Taft et al., 2020). Therefore, most studies focus on students (58.25%), teachers (25.24%), employees (8.74%), and believers of different religions (7.77%; Sakallı et al., 2021).

However, tolerance/intolerance, as a personality trait, can manifest itself as early as the age of 4, so its development depends primarily on family upbringing (Miklikowska, Thijs & Hjerm, 2019; Odenweller & Harris, 2018; Verkuyten & Killen, 2021). Since one's life and familiarity with diversity begin in the family, one should note the importance of tolerance for mutual understanding between close relatives, especially in parent-child relationships. Family upbringing is crucial for the formation of tolerance in children, who, once they become parents themselves, could teach it to their children. However, the nurturing of tolerance in children (within their families) is poorly studied.

Tolerance in parent-child relationships is of particular importance in collectivistic cultures, where awareness of oneself as “we” and maintaining relationships with loved ones, primarily family, is extremely important, as it forms the basis of a person’s emotional and personal well-being. Many psychological phenomena are associated with this dimension of culture. Collectivism-individualism affects the experiences of representatives of these cultures (Namcoong et al., 2021), and there is evidence of neurophysiological differences in the perception of reality among collectivists and individualists (Ng et al., 2010).

Thus, increased attention to cross-cultural research can expand the possibilities of social psychology (Triandis, 1983) and become the key to the success of the development and implementation of psychological, socio-psychological, and pedagogical programs for the formation of tolerance,

taking into account its significance for different generations, different social, gender and other groups.

In order to identify the significance of tolerance in parent-child relationships and between social groups within the same culture, the article uses the analysis of large international data which has become widespread not only in sociology but also in other sciences (Devellennes & Loveless, 2022; Hildebrandt et al., 2018; Ryumshina, Zinchenko et al., 2022). The advantages of such an analysis lie in the fact that it allows taking into account the representations of a particular psychological phenomenon by a large number of respondents with different socio-demographic characteristics, as well as identifying differences between them, including sociocultural ones. Thus, the conducted study is interdisciplinary in nature, combining the ideas of psychological science with rigorous methods of quantitative comparative analysis.

China in an era of change

China may be the most convenient model for analyzing this topic because, although it has regional differences in collectivism (Ren et al., 2021), it is a type of collectivist culture characteristic of many Asian (e.g., Japan, Korea, Uzbekistan, Tajikistan) and non-Asian countries (e.g., Russia).

The culture of Chinese society is family-oriented (Yang & Zhang, 2020). At the same time, the formation of the Chinese social structure is significantly influenced by Confucian culture, where the family is seen as the ideal embodiment of social relations, and the emotional interaction between its members contributes to harmony and well-being within it (Chuang, 2005). Relationships between parents and children are more significant than the relationships between husband and wife and between brothers and sisters, and intergenerational relationships are of great importance as well (Goh & Kuczynski, 2009; Yang et al., 2020).

As in many other collectivist cultures, Chinese children, growing up, are still functionally dependent on their parents. Thus, family members at all stages of life play a special role in the life of the Chinese. Various studies indicate the importance of parental support for the younger generation (Yang et al., 2020). For example, adolescents' anxiety, when they find themselves

faced with serious stressful situations, decreases if their relationship with their parents is close and increases if it is not (Yang & Yeh, 2006).

The second important reason for choosing China for the analysis of the stated topic is related to the fact that China has experienced significant economic growth associated with industrialization, urbanization, and globalization over the past few decades. Naturally, this led to serious social changes, which were also facilitated by the "one child" policy, the expansion of education, and the widespread use of the Internet (Li, 2020). The gap in socio-economic conditions led to the emergence of generations of the 80s, 90s, and 2000s, whose personal characteristics were formed in a different social environment; therefore, it is reasonable to assume that their views, values, interests, behavior, etc., differ from previous generations, which reasonably causes concern for scientists (Li, 2020).

The "one child" policy has led to increased focus on children. They became so valuable that the older members of the family began to subordinate their own interests, aspirations, and desires to the only child. Thus, the children began to exert a dominant position between generations (Yang et al., 2020), and, at the same time, the authority of the elder in the family was called into question (Zhou, 2001). The styles of upbringing in Chinese families and traditional family relations between generations have changed - from respect for the elderly to care for the youth (Yang et al., 2020).

The new generation grew up in a more comfortable economic climate and a more tolerant and liberal social environment. It is more open, characterized by independent views, self-confidence, and greater willingness to participate in public and political affairs compared to previous generations, and it becomes the main force that plays an important role in the main social transformations of the country (Li, 2020; Zhou, 2016). However, the new generation is not homogeneous; its socio-economic stratification is observed in various aspects of public life, and there is a significant difference between urban and rural youth of the second generation (Zhang et al., 2003). According to a number of researchers, this intragenerational stratification is no less significant than the stratification between generations (Li, 2020).

According to some researchers, relationships between generations are becoming more equal and close, and relationships between parents and children are more democratic (Xiao, 2016; Yang et al., 2020). However, it

should be noted that, due to cultural traditions, a different educational environment is created for boys and girls in Chinese families. The upbringing of a boy takes place in conditions of rigorous requirements for behavior, as he is the bearer of the family name, traditions, and customs. Girls are the home keepers. Therefore, parents pay more attention to their psychological education. A girl should be tolerant, respectful, and sociable. At the same time, filial piety, the central concept of Confucianism, is deeply rooted in Chinese culture and the subjective consciousness of the Chinese, giving them a sense of subjective security and belonging to their culture (Zheng & Li, 2022). Thus, mutual filial piety, as a high-quality interaction between children and parents, can enhance young people's ability to control and use their emotions, which, in turn, increases their life satisfaction (Chen et al., 2018) and reduces psychological stress (Wu & Chen, 2020). In other words, filial piety is still central to intergenerational bonds and commitments and continues to be an important cultural ideal that defines intergenerational caregiving responsibilities for many Chinese families.

At the same time, we can talk about the contradictory conclusions of scientists regarding whether filial piety is useful or harmful for individual development (Yeh & Bedford, 2003), and some studies show that faith in filial piety is weakening (Zhang et al., 2019). In order to prevent conflicts between parents and the younger generation and gain an understanding of how beliefs in filial piety affect the life satisfaction of young people and the loneliness of their parents (Chen et al., 2018), a more thorough study of this aspect of relationships in Chinese families is necessary (Zhang et al., 2019).

Thus, the combination of tradition, modernization, and the infiltration of Western individualism has resulted in a complex social psychology of Chinese youth. Although youth have become more individualistic, collectivism is still a stronger predictor of their values (Weng et al., 2021). In any case, interpersonal tolerance is extremely important for the mutual understanding between generations whose personal characteristics were formed under the influence of various social values. A rather revealing study in this respect, conducted in 2019 (Cheung et al., 2019), showed that adolescents and their parents and the parents of their parents all have different ideas of harmony in relationships with each other.

According to the World Values Survey, organized by Inglehart on the basis of his theory of socio-cultural changes in the conditions of modernization, we can talk about two post-modernization shifts: from traditional to industrial and from industrial to post-industrial (Inglehart, 1997). These shifts often lead to dramatic conflicts between generations of people who have been socialized in different social settings (traditional and transitional societies). The industrialization of society, which gives more opportunities to meet basic needs, leads to the formation of secular-rational values, in connection with which authority, order, and security are highly valued. The post-modernization shift in values led to the formation of post-materialistic values, such as tolerance, self-knowledge, self-expression, and increased civic activity. In Inglehart's theory, along with a change of generations and changes in values, the degree of tolerance in a society serves as an important indicator for measuring the transition from a traditional value system to a modern one (Inglehart, 1997). Thus, social tolerance (i.e., tolerance to certain social phenomena, such as behaviors and lifestyles of other individuals and groups that are legal and do not harm the interests of other people) is a kind of criterion for this transition.

Will the values of generations that grew up in China in different socio-economic conditions coincide? What place will tolerance take among these values, and for which social groups is it more important? Will the importance of tolerance for new generations grow with globalization, the growth of economic well-being, and internationalization, taking into account their socio-economic stratification? These are the questions that this empirical study aims to answer.

Thus, the aim of the present study was to examine the importance of nurturing tolerance in children within Chinese families. To ensure representativeness, we studied the views of adult respondents, who belong to different generations and live in different socio-economic conditions.

Method

The study used big data from the World Values Survey. The World Values Survey (WVS) questionnaire includes questions on various aspects of human life (religion, politics, participation in public organizations, gender relations, subjective well-being, etc.), as well as an extensive survey related

to socio-demographic data. Depending on the global geopolitical situation, different waves of data collection may include some additional questions. However, some questions are constant across waves, for example, questions about family and family upbringing.

The study used statistical analysis of data from the World Values Survey, in which China participated six times: in 1990 (second wave), 1995 (third wave), 2001 (fourth wave), 2007 (fifth wave), 2012/13 (sixth wave), 2018 (seventh wave) (Inglehart et al., 1990, 1995, 2001, 2007, 2012/13, 2018). As for the number of respondents, 1000 people took part in the second wave, 1500 in the third, 1000 in the fourth, 1991 in the fifth, 2299 in the sixth, and 6699 in the seventh. Thus, our sample included 14489 people with different educational backgrounds and social classes, aged 16 to 50 and over.

The following statistical methods were used for data analysis: frequency analysis, logistic regression, Kruskal-Wallis test, Conover test, and confidence intervals. All calculations were performed using the R programming language R within RStudio interface (Kupriyanov & Yavna, 2016).

Results

To address our research aim, we selected the answers of Chinese respondents from various social groups regarding the significance of nurturing certain personal qualities within the family. The question was formulated as follows: "In front of you on the card, there is a list of qualities that can be nurtured in children in the family. Which of them, if any, do you think are the most important?". Answering this question, respondents could choose 5 qualities out of 11, including tolerance (Table 1).

Table 1

Important Qualities for Upbringing in the Family (%)

Qualities	2nd wave (1990)	3rd wave (1995)	4th wave (2001)	5th wave (2006)	6th wave (2011)	7th wave (2017)
Good manners	47.3	66.8	NA	NA	NA	83.7
Independence	84.0	50.1	74.1	69.8	69.7	78.2

Diligence	64.6	72.7	85.8	83.4	75.3	70.4
Responsibility	66.9	34.5	63.6	67.2	65.9	78.4
Imagination	26.7	22.1	35.3	25.3	17.0	21.6
Tolerance	61.7	43.0	72.6	64.7	52.2	60.5
Thrift	55.6	62.2	57.2	61.6	50.7	40.3
Determination, perseverance	45.0	36.3	16.0	24.0	26.0	20.5
Religious faith	1.2	3.3	NA	2.4	1.2	1.1
Unselfishness	30.9	28.2	37.0	30.8	29.2	28.7
Obedience	8.5	29.3	14.8	13.8	7.5	5.7

As one can see, among the distinguished qualities important for family upbringing, tolerance is a fairly significant quality, taking the 2nd-5th rank, depending on the wave (i.e., year). The qualities consistently assessed as more significant for the upbringing of the younger generation than tolerance are independence, diligence, and, possibly, good manners (data on good manners are not available in all waves, preventing us from drawing any definitive conclusions).

When looking into the significance of nurturing tolerance within each wave separately, the wave in which the highest percentage of respondents ranked confidence as important was the fourth (2001); in the third wave (1995), the lowest percentage ranked it as important (Table 2).

Table 2

The Significance Tolerance across Waves (%)

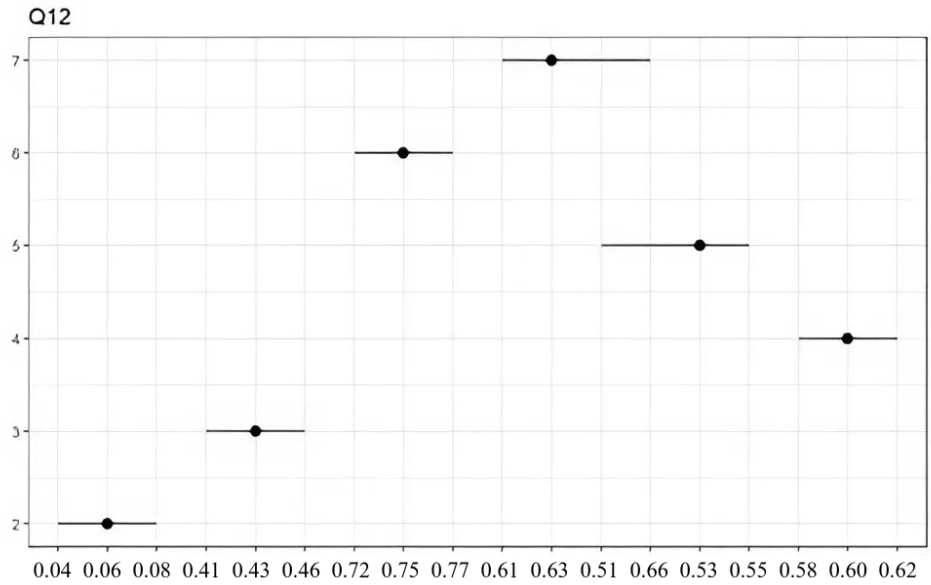
Wave	Important child qualities: tolerance and respect for other people	
	Important	Not mentioned
The seventh wave, 2018	60.5%	39.0%
The sixth wave, 2012/13	52.2%	47.8%
The fifth wave, 2007	64.7%	35.3%

The fourth wave, 2001	72.6%	27.03%
The third wave, 1995	43.0%	57.0%
The second wave, 1990	61.7%	38.3%

The results of the logistic regression confirm the previously presented data (Figure 1).

Figure 1

The Dynamics of Tolerance Across Waves (Logistic Regression)

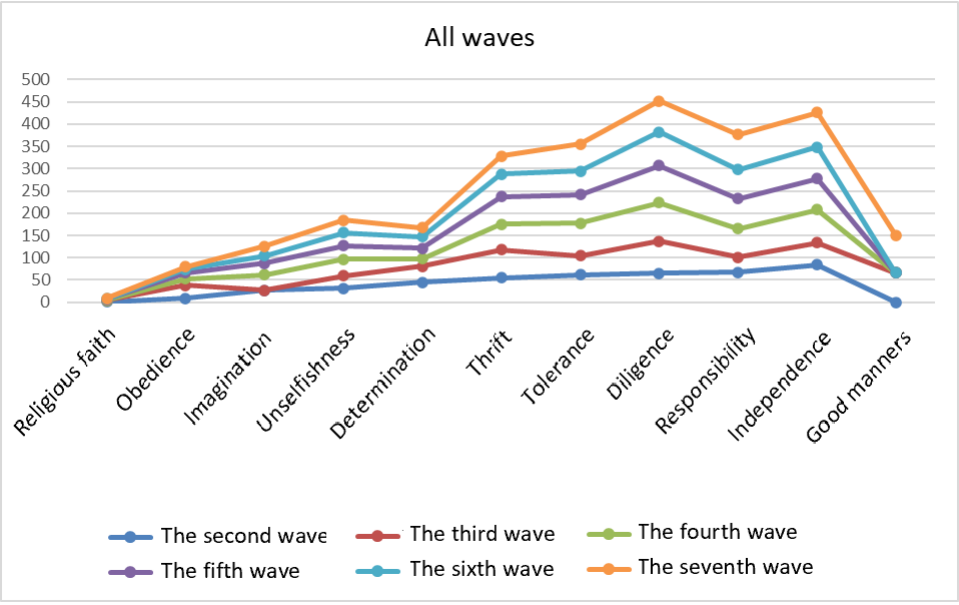


A comparative analysis of the sample (N = 10827) was conducted using the Kruskal-Wallis test. Six waves act as independent samples; the wave of the study, which allowed us to identify year-to-year changes in variables (i.e., tolerance, gender, age, level of education, social class), was chosen as a grouping variable. According to the results, most variables show a high level of reliability of changes ($p < 0.001$) with an insignificant effect size. Then, using the Conover test, the direction of changes was determined, and a tendency towards a decrease in the level of tolerance was identified ($p < 0.001$). We note that, in 1995, there was a sharp decrease in the significance

of the development of several qualities that were important for respondents who took the survey in other years.. In addition to tolerance, these qualities include responsibility, independence, and unselfishness. At the same time, the significance of not only religious devoutness, but also obedience, frugality, and diligence sharply increased (Figure 2).

Figure 2

Qualities Important to be Developed within a Family (Waves 2-7)



The distribution of the significance of qualities important for development in different age groups is presented in Table 3. Respondents under the age of 29 in the second wave had the highest endorsement of independence (89.8%) - higher than the endorsement of this quality in other age groups and higher than the endorsement of any other qualities in any other wave. Diligence was found important to participants in all waves and age groups and most important for respondents in the fourth wave, particularly those aged 30-49. Responsibility (78.9%) was particularly highly endorsed by respondents in the seventh wave – relatively equally across age groups The highest value of tolerance was observed in the fourth wave,

among respondents under the age of 29 and respondents aged 30 to 49. If we look at the younger generation, it was only in the second wave that they appreciated tolerance to a somewhat lesser extent than older age groups. From the third wave onwards, the importance of tolerance development for respondents under 30 exceeded its importance for older generations. As for the respondents aged 30-49, starting from the fourth wave, the significance of tolerance either exceeded or was equal to its significance among persons aged 50 and over.

Table 3

The Importance of Particular Qualities across Age Groups

Qualities		2nd wave (1990)	3rd wave (1995)	4th wave (2001)	5th wave (2006)	6th wave (2011)	7th wave (2017)
Independence	up to 29 y.o.	89.8% (*N* = 303)	53.2% (*N* = 432)	78.4% (*N* = 194)	76.8% (*N* = 564)	75.1% (*N* = 501)	84.1% (*N* = 679)
	30-49 y.o.	83.1% (*N* = 443)	49.4% (*N* = 755)	77.1% (*N* = 572)	72.8% (*N* = 895)	71.3% (*N* = 1,063)	81.0% (*N* = 1,348)
	50 y.o. and more	78.7% (*N* = 254)	47.6% (*N* = 313)	63.2% (*N* = 234)	57.3% (*N* = 532)	63.8% (*N* = 735)	70.6% (*N* = 1,010)
Diligence	up to 29 y.o.	51.2% (*N* = 303)	67.6% (*N* = 432)	80.9% (*N* = 194)	80.9% (*N* = 564)	68.8% (*N* = 501)	62.1% (*N* = 679)
	30-49 y.o.	65.2% (*N* = 443)	74.7% (*N* = 755)	89.2% (*N* = 572)	86.8% (*N* = 895)	77.3% (*N* = 1,063)	71.1% (*N* = 1,348)
	50 y.o. and more	79.5% (*N* = 254)	74.8% (*N* = 313)	81.6% (*N* = 234)	80.4% (*N* = 532)	76.8% (*N* = 735)	75.1% (*N* = 1,010)
Responsibility	up to 29 y.o.	67.0% (*N* = 303)	32.6% (*N* = 432)	72.2% (*N* = 194)	73.7% (*N* = 564)	66.0% (*N* = 501)	78.9% (*N* = 679)
	30-49 y.o.	65.5% (*N* = 443)	33.9% (*N* = 755)	62.6% (*N* = 572)	67.9% (*N* = 895)	65.8% (*N* = 1,063)	79.7% (*N* = 1,348)
	50 y.o. and more	69.3% (*N* = 254)	38.3% (*N* = 313)	59.0% (*N* = 234)	58.9% (*N* = 532)	66.0% (*N* = 735)	76.3% (*N* = 1,010)
Tolerance	up to 29 y.o.	58.4% (*N* = 303)	49.1% (*N* = 432)	74.7% (*N* = 194)	73.7% (*N* = 564)	56.3% (*N* = 501)	65.6% (*N* = 679)

30-49 y.o.	63.2% (*N* = 443)	39.9% (*N* = 755)	74.3% (*N* = 572)	65.5% (*N* = 895)	51.0% (*N* = 1,063)	63.3% (*N* = 1,348)
50 y.o. and more	63.0% (*N* = 254)	42.2% (*N* = 313)	66.7% (*N* = 234)	53.8% (*N* = 532)	51.1% (*N* = 735)	53.4% (*N* = 1,010)

Note. *N* is the total number of respondents of a particular age, participating in a particular wave.

According to logistic regression data, in the third wave (1995), the group under 29 had a higher tolerance than those aged 30–49, and they did not differ significantly from those aged 50 and over (Figure 3). There were no significant differences between age groups in terms of the significance of nurturing tolerance in the fourth wave (Figure 4).

Figure 3

Differences in the Importance of Tolerance among Different Age Groups in the 3rd Wave (1995)

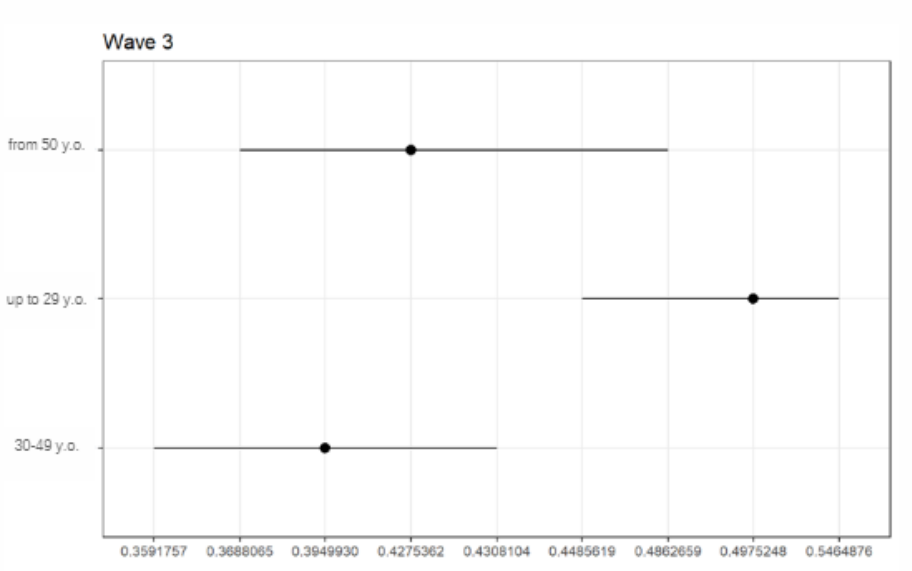
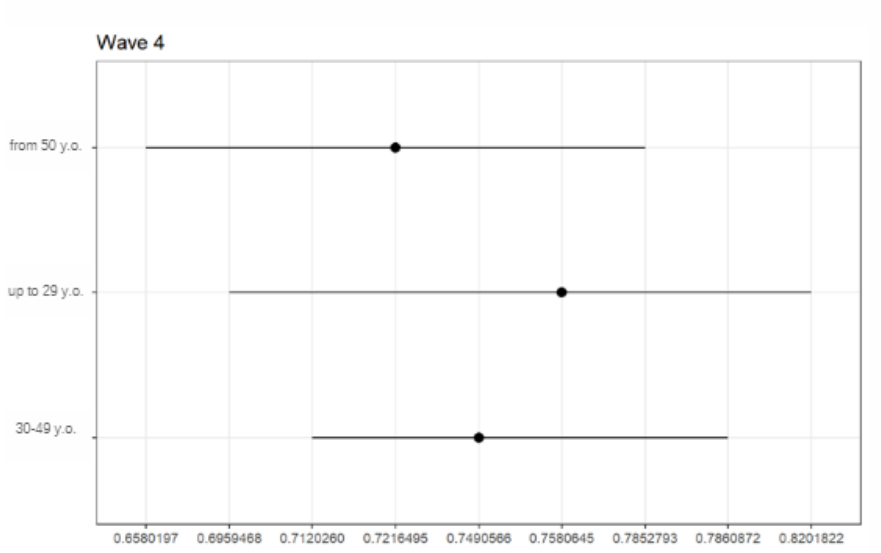


Figure 4

Differences in the Importance of Tolerance among Different Age Groups in the 4th Wave (2001)



In the fifth wave, tolerance was most significant among those under 29, followed by the group aged 30-49; it was the least significant in the group aged 50 and over (Figure 5). In the sixth wave, no significant differences were found (Figure 6).

Figure 5

Differences in the Importance of Tolerance among Different Age Groups in the 5th Wave (2007)

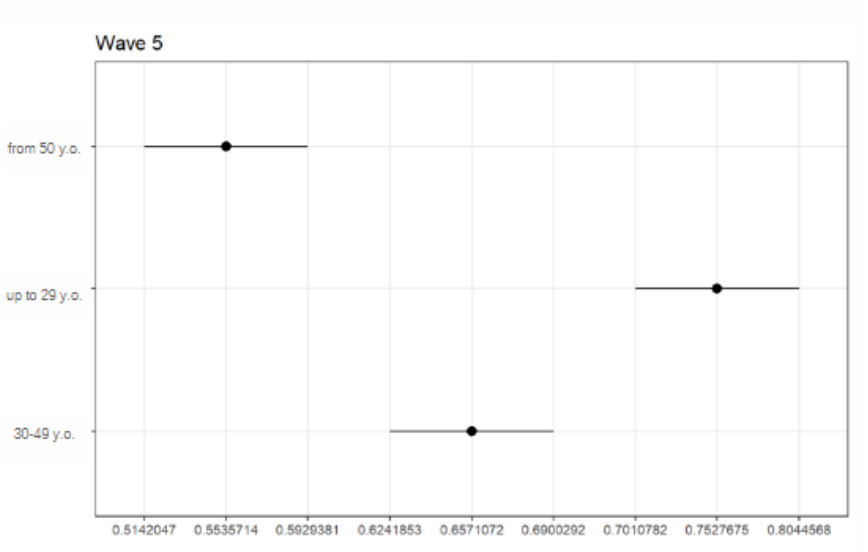
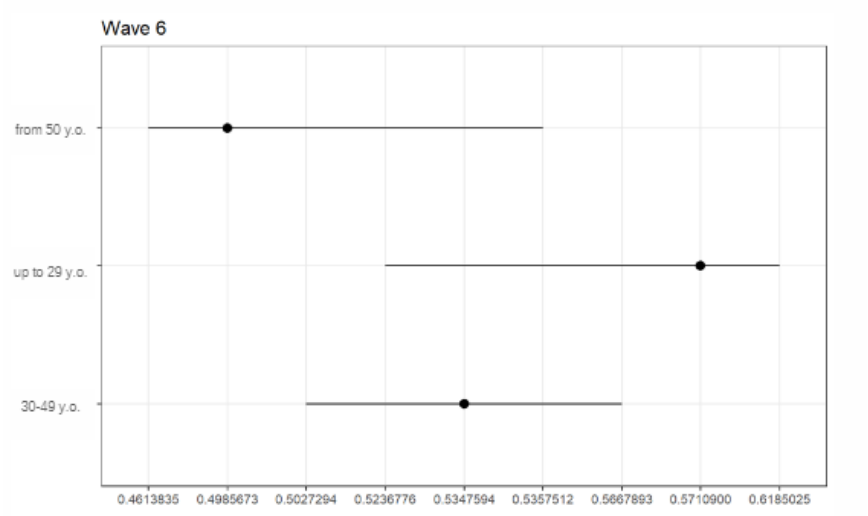


Figure 6

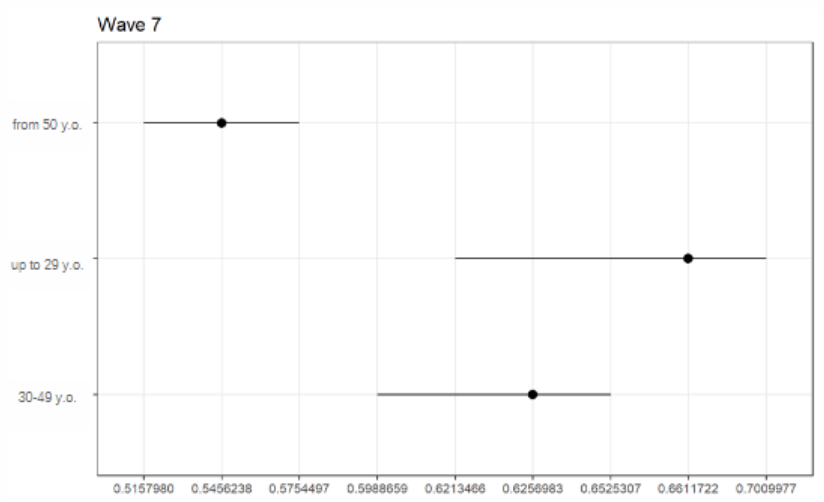
Differences in the Importance of Tolerance among Different Age Groups in the 6th Wave (2012/13)



In the seventh wave, there were no significant differences between the groups under 29 and aged 30-49; respondents aged 50 and over had the lowest significance of tolerance (Figure 7).

Figure 7

Differences in the Importance of Tolerance among Different Age Groups in the 7th Wave (2018)

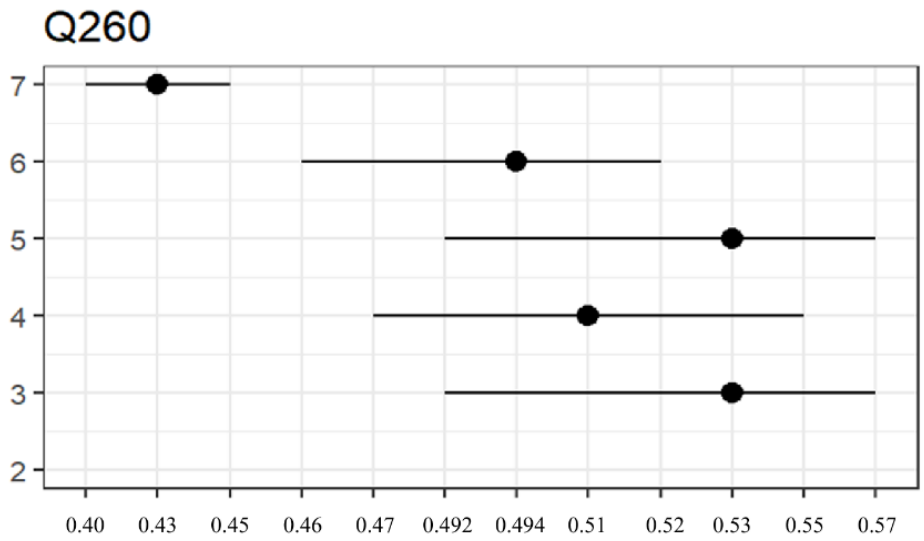


The work also reviewed statistically (using the confidence interval) the issue of socio-demographic indicators of respondents who consider it important to develop tolerance in the family. All waves were used, except for the second (1990), for which there were no accurate data.

The number of men from the third to the sixth waves (1995, 2001, 2007, 2012/13) fluctuated statistically insignificantly; in the seventh wave (2018), it decreased (Figure 8).

Figure 8

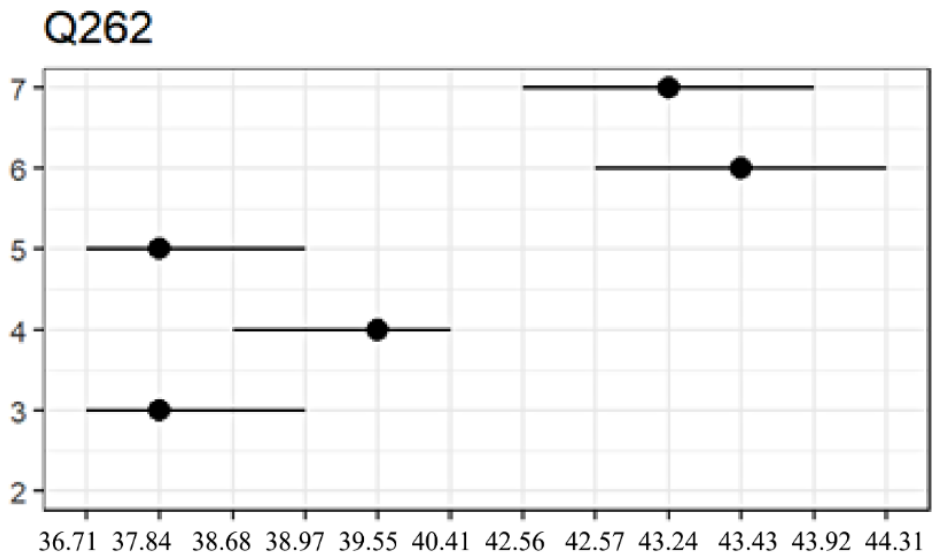
Gender Differences among Respondents in 1995 (Third Wave), 2001 (Fourth Wave), 2007 (Fifth Wave), 2012/13 (Sixth Wave), 2018 (Seventh Wave)



In the group of respondents aged 30-49, from the third to the fifth wave (from 1995 to 2007), there were statistically insignificant variations (37.8-39.5 when rounded, aged 38-40); in the sixth (2012/13) and the seventh wave (2018), the age of respondents increased (43 years old; Figure 9).

Figure 9

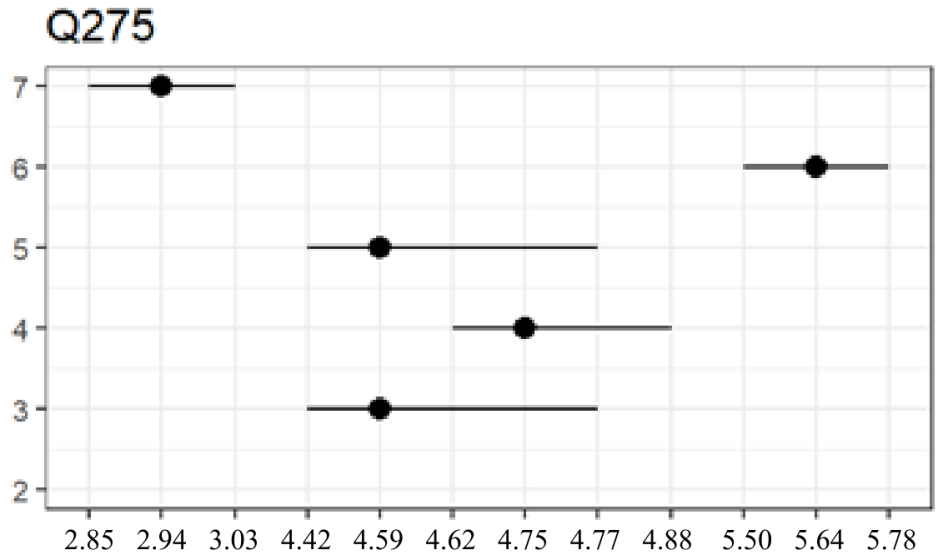
Age Differences Among Respondents in 1995 (Third Wave), 2001 (Fourth Wave), 2007 (Fifth Wave), 2012/13 (Sixth Wave), 2018 (Seventh Wave)



The level of education of respondents also did not change significantly from the third to the fifth wave; in the sixth (2012/13), it increased; in the seventh (2018), it became lower than it had been in the earlier waves. Thus, the level of education among respondents who considered it important to develop tolerance has become lower (Figure 10).

Figure 10

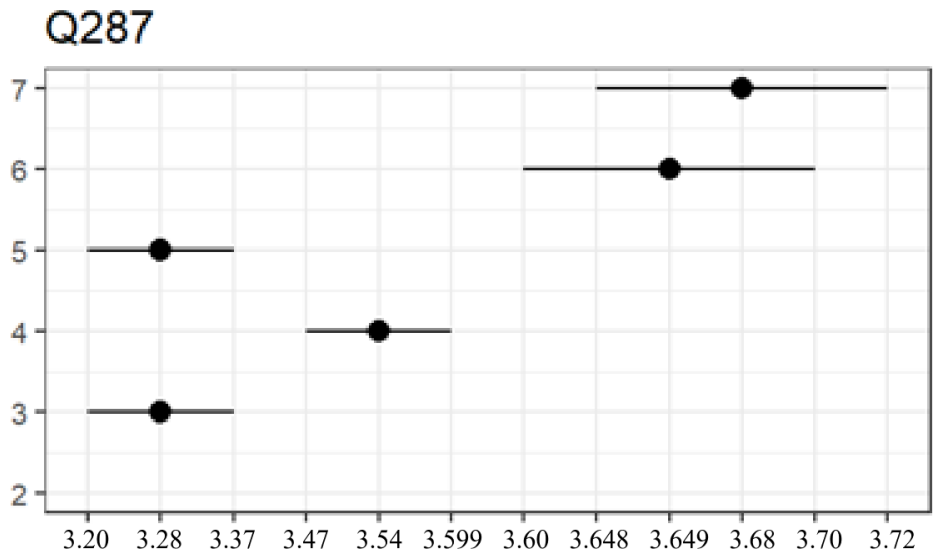
Differences in the Level of Education among Respondents in 1995 (Third Wave), 2001 (Fourth Wave), 2007 (Fifth Wave), 2012/13 (Sixth Wave), 2018 (Seventh Wave)



The indicator of the social class of respondents in the fourth wave (2001) increased compared to the third (1995); in the fifth (2007), it returned to its previous level. However, in the sixth (2012/13) and seventh (2018) waves, it increased significantly (Figure 11).

Figure 11

Differences in Social Class among Respondents in 1995 (Third Wave), 2001 (Fourth Wave), 2007 (Fifth Wave), 2012/13 (Sixth Wave), 2018 (Seventh Wave)



Discussion

As Hofstede wrote, a researcher who analyzes cultural dimensions, to a certain extent, acts as a representative of his culture (Hofstede, 2007). Agreeing with this, we note that the purpose of the article was not a thorough analysis of the relationship between Chinese parents and children, as well as generations. Using the example of this unique ancient culture, the authors wanted to show the possibility and significance of using big data while analyzing and forecasting intergenerational relationships in a period of great change.

The inclusion of tolerance in the list of important qualities for the younger generations indicates that it is of great value (Devellennes & Loveless, 2022; Inglehart, 1997). At present, for many collectivist cultures in the East (e.g., countries that used to be part of the USSR), experiencing a

rapid economic and political boom, it is important to study tolerance between parents and children, as well as generations that grew up in the context of different social values more generally. Some researchers studying tolerance take into account not only age, social status, education, and place of birth, but also the number of children and the order of birth (Abidova, 2016).

Is tolerance considered important in modern China? According to the opinion of Chinese respondents in our study, the younger generation needs to develop, first of all, diligence, and independence. "Diligence" - the most significant quality since 1995- lost its first place by the seventh wave, giving way to "independence" and "responsibility." Nevertheless, it must be assumed that “independence” and “diligence” (rather than tolerance) are the “cementing” values of generations, which is consistent with studies of Chinese proverbs, in which a large place is given to “diligence,” “free spirit” (Weng et al., 2021), and dedication (Yue & Ng, 1999). Unfortunately, not all the waves had data on the significance of the development of “good manners.” Therefore, one can only assume that this quality is also considered important, as it was in the seventh wave, where it came to the fore. Tolerance, which can promote mutual understanding in parent-child relationships and relationships between generations, was also important for respondents, although not as important as diligence and independence. Of greatest interest are the respondents participating in the survey after 2001 (waves 5-7), who represent the new generations of the 1980s, 1990s, and 2000s. Our results showed that tolerance is indeed highly valued by these generations (especially those born in the 1980s), which is in line with a number of previous studies and the CSS data (Li, 2020).

Of course, the socio-economic environment affects the importance of developing tolerance in the family (Li, 2020; Zhang et al., 2003; Zhou, 2016). This can explain the surge in its significance in the wave of 2001, where the highest indicator of the significance of tolerance was recorded among respondents under 29. However, starting from 1995 and on, the importance of tolerance development for young respondents (under 30), although not a leading value, exceeded the importance of this quality for older generations. Importance of tolerance may also be related to age (i.e., young people are usually more tolerant than older generations). Moreover, the respondents

who took part in the survey in 1990 reported approximately the same importance of developing tolerance within the family as those who participated in the seventh wave (2018) - a finding that requires additional research. Similarly, while remaining a fairly important quality, in 2011, the significance of tolerance decreased.

By 2018, the social class among respondents who consider it important to develop tolerance in the family significantly increased, but their level of education became lower. Perhaps this can be explained by the number of men among the respondents in this wave (which has decreased), as women are less educated than men. The question of filial piety also remains open, although the fears of some Chinese scholars about its change, in our opinion, are not groundless.

Limitations and future directions

The lack of data on the place of residence prevented us from considering regional variations and potential differences in the significance of nurturing tolerance between rural and urban residents. Due to incorrect socio-demographic characteristics in the second wave and the lack of data on gender and the importance of nurturing good manners and piety, it was not possible to fully compare the importance of nurturing these qualities versus nurturing tolerance within the family. Thus, these limitations of the present study may serve as prospects for future research. Further research may also assess whether and how a broader set of socio-economic and socio-psychological factors influence the importance of nurturing tolerance within the family.

Conclusion

According to the results, tolerance is quite important for Chinese respondents; still, its significance varied depending on the time the survey was administered (i.e., the wave) and was lower than the significance of "diligence" and "independence." The increasing importance of nurturing "independence," "responsibility," and "good manners" in the family by 2018 is the most striking manifestation of the influence of the new socio-economic situation on the representation of the qualities that the younger generations should possess. In line with this is that there were more respondents who

considered it important to form tolerance in the family in the generation of the 1980s than among other generations. The highest indicator of the significance of tolerance was recorded in 2001 among respondents under 29. Thus, the obtained results allow us to conclude that the importance of nurturing tolerance in the family will differ depending on the age of the respondents. New socio-economic conditions have led to the formation of a more tolerant generation.

Among other socio-demographic characteristics, one can note the influence of social class on the assessment of the significance of nurturing tolerance within the family (the higher the social class, the more significant the tolerance is) and the level of education of the respondents. The level of education is lower in 2018 compared to earlier years. Finally, several questions require further research, including whether there are gender differences in the significance of tolerance for Chinese respondents.

Conflict of interest

We have no conflicts of interest to disclose.

Data availability statement

The data used for this study are available from the World Values Survey (WVS) at https://www.worldvaluessurvey.org/WVSContents.jsp

References

Abidova, Z. (2016). *The essence and realities of the phenomenon of "tolerance"*. Uzbekistan, Tashkent, Navro'z.

Armstrong, G., & Wronski, J. (2019). Framing Hate: Moral Foundations, Party Cues, and (In)Tolerance of Offensive Speech. *Journal of Social and Political Psychology*, *7*(2), 695-725. https://doi.org/10.5964/jspp.v7i2.1006

Bayram Özdemir, S., & Özdemir, M. (2020). The Role of Perceived Inter-Ethnic Classroom Climate in Adolescents' Engagement in Ethnic Victimization: For Whom Does it Work?. *Journal of youth and adolescence, 49*(6), 1328–1340. https://doi.org/10.1007/s10964-020-01228-8

Brenik, A., Lawrence, J., & Berger, R. (2019). Teach tolerance or behave tolerably? Assessment of measures to reduce prejudice based on skills and contacts among Palestinian-Israeli and Jewish-Israeli youth. *Journal of School Psychology*, *75*, 8-26. https://doi.org/10.1016/j.jsp.2019.07.001

Chen, W., Yan, J. & Chen, C. (2018). Lesson of Emotions in the Family: The Role of Emotional Intelligence in the Relation Between Filial Piety and Life Satisfaction Among Taiwanese College Students. *Asian Journal of Social Psychology*, *21*, 74-82. https://doi.org/10.1111/ajsp.12207

Cheung, F., Lun, V., Ngo, H. & Fong, E. (2019). Seeking harmony in Chinese families: A dyadic analysis on Chinese parent–child relations. *Asian Journal of Social Psychology*, *23*, 82-90. https://doi.org/10.1111/ajsp.12393

Chuang, Y. (2005). Effects of interaction pattern on family harmony and well-being: Test of interpersonal theory, Relational-Models theory, and Confucian ethics. *Asian Journal of Social Psychology*, *8*(3), 272-291. https://doi.org/10.1111/j.1467-839X.2005.00174.x

Devellennes, C., & Loveless, P. M. (2022). The tolerance of the despised: Atheists, the non- religious, and the value of pluralism. *International Political Science Review*, *43*(4), 580-594. https://doi.org/10.1177/01925121211034148

Goh, E. C. L., & Kuczynski, L. (2009). Agency and Power of Single Children in Multi-Generational Families in Urban Xiamen, China. *Culture & Psychology*, *15*(4), 506–532. https://doi.org/10.1177/1354067X09344881

Hamer, K., McFarland, S. & Penczek, M. (2019). What lies beneath? Predictors of identification with all humanity. *Personality and Individual Differences*, *141*(2), 258-267. https://doi.org/10.1016/j.paid.2018.12.019

Hildebrandt, A., Trüdinger, E.-M. & Wyss, D. (2018). The Missing Link? Modernization, Tolerance, and Legislation on Homosexuality. *Political Research Quarterly*, *72*(3), 539-553. https://doi.org/10.1177/1065912918797464

Hofstede, G. (2007). A European in Asia. *Asian Journal of Social Psychology*, *10*(1), 16-21. https://doi.org/10.1111/j.1467-839X.2006.00206.x

Inglehart, R. (1997). *Modernization and postmodernization: cultural, economic, and political change in 43 societies*. Princeton, New Jersey: Princeton University Press.

Inglehart, R. et al. (1990, 1995, 2001, 2007, 2012/13, 2018). *World Values Survey* [Data set]. Retrieved April 10, 2023 from: https://www.worldvaluessurvey.org/WVSContents.jsp

Jung, Y., Hall, J., Hong, R., Go, T., Ong, N. & Tan, N. (2014). Payback: the influence of relationships and cultural norms on reciprocity. *Asian Journal of Social Psychology, 17*(3), 160-172. https://doi.org/10.1111/ajsp.12057

Kupriyanov I.V., Yavna D.V. (2016). *Introduction to the use of the statistical processing system R in psychological research*. Russia, Moscow, KREDO.

Li, C. (2020). Children of the reform and opening-up: China's new generation and new era of development. *The Journal of Chinese Sociology*, *7*(18), 1-22. https://doi.org/10.1186/s40711-020-00130-x

Miklikowska, M., Bohman, A., & Titzmann, P. F. (2019). Driven by context? The interrelated effects of parents, peers, classrooms on development of prejudice among Swedish majority adolescents. *Developmental Psychology*, *55*(11), 2451–2463. https://doi.org/10.1037/dev0000809

Miklikowska, M., Thijs, J. & Hjerm, M. (2019). The Impact of Perceived Teacher Support on Anti-Immigrant Attitudes from Early to Late Adolescence. *Journal of Youth Adolescence*, *48*, 1175–1189. https://doi.org/10.1007/s10964-019-00990-8

Namcoong J., Ro J., Hur T. & Chey Y. (2021). Regrets about "I" versus regrets about "us": the influence of culture on regrets about actions versus inaction. *Asian Journal of Social Psychology*, *25*(3), 532-543. https://doi.org/10.1111/ajsp.12515

Ng, S. H., Han, S., Mao, L. & Lai, K. L. J. (2010). Dynamic Bicultural brain: An MRI study of their flexible neural representation of themselves and significant others in response to cultural primes. *Asian Journal of Social Psychology*, *13*(2), 83-91. https://doi.org/10.1111/j.1467-839X.2010.01303.x

Odenweller, K. & Harris, T. (2018). Intergroup Socialization: The Influence of Parents' Family Communication Patterns on Adult Children's Racial Prejudice and Tolerance. *Communication Quarterly*, *66*(5), 501-521. https://doi.org/10.1080/01463373.2018.1452766

Park, J., Berry, J. & Joshanloo, M.(2022). Japanese people's attitudes toward acculturation and intercultural relations. *Journal of Pacific Rim Psychology*, *16*(54), 1-15. https://doi.org/10.1177/18344909221090996

Parkes, L., Schneider, S. & Bochner, S. (2002). Individualism-collectivism and self-concept: Social or contextual? *Asian Journal of Social Psychology*, *2*(3), 367-383. https://doi.org/10.1111/1467-839X.00046

Ren, X., Cang, X. & Ryder, A. (2021). An Integrated Ecological Approach to Mapping Variations in Collectivism Within China: Introducing the Triple-Line

Framework. *Journal of Pacific Rim Psychology*, *15*, 1-12. https://doi.org/10.1177/1834490921991436

Rjumshina, L. (2000). Empirical study of pedagogical activity styles. *Voprosy Psikhologii*, *1*, 148-149.

Ryumshina, L., Berdyanskaya, Y., Altan-Avdar, I. & Belousova, A. (2022). Symmetrical and asymmetrical approaches to communication in education in distance learning. *Lecture Notes in Networks and Systems*, *247*, 471–481. https://doi.org/10.1007/978-3-030-80946-1_45.

Ryumshina, L., Zinchenko, E., Chernova, A., Berdyanskaya, J. & Batychko, O. (2022). Dynamics of Attitudes towards Tolerance Formation in the Family of Various Social Groups Representatives. *Russian Psychological Journal*, *19*(4), 211–230. https://doi.org/10.21702/rpj.2022.4.14

Sakallı, Ö., Tlili, A., Altınay, F., Karaatmaca, C., Altınay, Z., & Dağlı, G. (2021). The Role of Tolerance Education in Diversity Management: A Cultural Historical Activity Theory Perspective. *SAGE Open*, *11*(4), 1-11. https://doi.org/10.1177/21582440211060831

Sandoval-Hernandez, A., Miranda, D. & Isac, M.M. (2018). How Do We Assess Civic Attitudes Toward Equal Rights? Data and Methodology. In: Sandoval-Hernández, A., Isac, M.M., Miranda, D. (eds), *Teaching Tolerance in a Globalized World*, *4*, 11-17. IEA Research for Education, Springer Open, Cham. https://doi.org/10.1007/978-3-319-78692-6_2

Shestakova, T., Shestakova, A., Belova, E., Zinchenko, E. & Fokin, N. (2022), Educational leaders training in the context of digitalization. *Lecture Notes in Networks and Systems*, *403*, 1075-1084. https://doi.org/10.1007/978-3-030-96383-5_120

Song, Y. (2018). The moral virtue of open-mindedness. Canadian Journal of Philosophy, 48(1), 65–84. https://doi.org/10.1080/00455091.2017.1335566

Taft, L.J., Woods, K. & Ford, A. (2020). The contribution of the Educational Psychology Service to community cohesion: A grateful inquiry. *Pedagogical psychology in practice, 36*(1), 1-16. https://doi.org/10.1080/02667363.2019.1650722

Triandis, H. C. (1983). Some dimensions of intercultural variation and their implications for community psychology. J*ournal of Community Psychology, 11*(4), 285-302. https://doi.org/10.1002/1520-6629(198310)11:4<285::AID-JCOP2290110403>3.0.CO;2-8

Verkuyten, M. & Killen, M. (2021). Tolerance, Dissenting Beliefs, and Cultural Diversity. *Child Development Perspectives*, *15*(1), 51-56. https://doi.org/10.1111/cdep.12399

Weng, L., Zhang, Y., Kulich, S. & Zuo, C. (2021). Cultural values in Chinese proverbs reported by Chinese college students. *Asian Journal Of Psychology*, *24*, 232-243. https://doi.org/10.1111/ajsp.12415

Wu, C. W. & Chen, W. (2020). Mediating role of regulatory focus in the relation between filial piety and youths' life satisfaction and psychological distress. *Asian Journal of Social Psychology, 24*(4), 499-510. https://doi.org/10.1111/ajsp.12447

Xiao, S. (2016). Intimate power: the intergenerational cooperation and conflicts in childrearing among urban families in contemporary China. *The Journal of Chinese Sociology*, *3*(18), 1-24. https://doi.org/10.1186/s40711-016-0037-y

Yang, Y.-J. & Yeh, K.-H. (2006). Differentiating the effects of enacted parental support on adolescent adjustment in Taiwan: Moderating role of relationship intimacy. *Asian Journal Of Social Psychology*, *9(2)*, 161-166. https://doi.org/10.1111/j.1467-839X.2006.00193.x

Yang, Y., Shi, Y. & Zhang, D. (2020). Intergenerational effects on individual charitable donation: an innovative study on philanthropy in China. *Chinese Journal of Sociology*, *22*, 1-23. https://doi.org/10.1186 / s40711-020-00139-2

Yeh, K. & Bedford, O. (2003). A test of the Dual Filial Piety model. A*sian Journal of Social Psychology*, *6*, 215-228. https://doi.org/10.1046/j.1467-839X.2003.00122.x

Yue, X. D. & Ng, S. H. (1999). Filial Obligations and Expectations in China: Current Views from Young and Old People in Beijing. *Asian Journal of Social Psychology, 2,* 215-226. https://doi.org/10.1111/1467-839X.00035

Zhang, X., Zheng, X. & Wang, L. (2003). A comparative study of the individual modernity of adolescents between urban and rural areas in China. *Asian Journal of Social Psychology*, *6*(1), 61-73. https://doi.org/10.1111/1467-839X.t01-1-00010

Zhang, M., Lin, T., Wang, D. & Jiao, W. (2019). Filial piety dilemma solutions in Chinese adult children: The role of contextual theme, filial piety beliefs, and generation. *Asian Journal of Social Psychology*, *23*(2), 227-237. https://doi.org/10.1111/ajsp.12395

Zheng, X. & Li, H. (2022). How Chinese children's filial piety beliefs affect their parents' life satisfaction and loneliness. *PsyCh Journal*, *11*(6), 863-872. https://doi.org/10.1002/pchj.558

Zhou, X. (2001). *Cultural Transmission: Parent-Child Relations in a Changing Society*. Retrieved January 21, 2023 from: https://www.researchgate.net/publication/237288401_Cultural_Transmission_Parent-Child_Relations_in_a_Changing_Society.

Zhou, X. (2016). Intergenerational gap in cultural return and media influence. *The Journal of Jiangsu Administration Institute*, *2,* 63–70.