Publishing
INASED

International Journal of Progressive Education

1554 - 5210

IJPE
INASED Published ©

ijpe.inased.org

IJPE International Journal of Progressive Education

Volume 22 Number 2 April 2026

An International Journal Published by International Association of Educators (INASED)

International Journal of Progressive Education

Frequency: Six times a year.

ISSN: 1554-5210

Owner & Publisher: International Association of Educators

Indexing/Abstracting:

1. **H.W. Wilson (Education Full Text)**: https://www.ebsco.com/m/ee/Marketing/titleLists/eft-coverage.htm
2. **EBSCO Publication**: http://www.ebsco.com
3. **Cabell's Directory of Publishing**: http://www.cabells.com
4. **Ulrichsweb:** http://ulrichsweb.serialssolutions.com/
5. **Open AIRE:** https://www.openaire.eu/

2026 Subscription Rates

- $120 Association Member USA (Canada: $130; Rest of World: $140)
- $140 Individual USA (Canada: $150; Rest of World: $170)
- $110 Student USA (Canada: $120; Rest of World: $140)
- $240 Library/Institution USA (Canada: $260; Rest of World: $360)
- Single Issues and Back Issues: $35 USA (Canada: $45; Rest of World: $55)

If you wish to subscribe for the printed edition of IJPE, please contact the Customer Service department at secretary@inased.org

TABLE OF CONTENTS

Volume 22, Number 2
April 2026

Articles

International Journal of Progressive Education, Volume 22 Number 2, 2026

Adaptation of the Dispositional Empathy with Nature Scale to Turkish: The Wolf Example

Filiz Bektaşlı*
Hacettepe University (Türkiye)

Sevilay Dervişoğlu**
Hacettepe University (Türkiye)

Abstract

Empathy towards nature and wild animals is becoming increasingly important in terms of environmental protection and sustainability. Measuring empathy towards nature is very important in this context. This article presents findings regarding the validity and reliability of Dispositional Empathy with Nature (DEN) Scale, adapted into Turkish to measure empathy towards wolves, in a sample of Turkish adolescents. The research was conducted with a total of 1140 secondary school students (M=15.72; SD=0.87), 618 female, 507 male and 15 gender-unspecified, studying in public schools in Ankara. As a result of exploratory factor analyses, it was seen that the internal consistency of the scale was high, and the single-factor structure was supported as in the original DEN scale. Females were found to have higher empathy for wolves than males.

Keywords: Wild animals, wolves, empathy, high school students

* ***Filiz Bektaşlı***, Graduate Student, Department of Mathematics and Science Education, Hacettepe University, e-mail: filizkocbektasli@gmail.com

** ***Prof. Dr. Sevilay Dervişoğlu***, Hacettepe University, Faculty of Education, Department of Mathematics and Science Education, ORCID ID: https://orcid.org/0000-0003-4302-6027 e-mail: sevilayd@hacettepe.edu.tr

Correspondence: filizkocbektasli@gmail.com

Introduction

Environmental crises, biodiversity losses, habitat destruction and excessive use of natural resources that occur frequently today are important issues that negatively affect both ecological and social well-being. For example, the extinction of many ecologically critical wild animal species leads to dramatic disruptions in essential ecosystem functions (Dirzo et al., 2014). Due to environmental concerns such as the decline of wild animal populations and the destruction of nature, researchers have turned to studies focusing on emotional variables such as empathy (Schultz, 2000; Berenguer, 2007; Ghasemi & Kyle, 2021; Sevillano et al., 2007). It is argued that empathy with nature is of critical importance in terms of re-establishing the connection that humans have lost with nature (Ives et al., 2018) and sustainability (Brown et al., 2019; Rifkin, 2009). Empathy is seen by environmental researchers as a key solution in efforts to protect nature and wildlife (Tam, 2013). This makes empathy with nature an important issue in environmental education (Schultz, 2000). The word *empathia*, which means "*feeling inside*" in Greek, was first defined by aesthetic theorists as "*the ability of an individual to perceive the emotional, intellectual and intuitive experience of another person*" (Goleman, 1998, p. 129). The concept of empathy has been defined from different perspectives since it was first introduced. In recent studies, 43 different definitions of the concept of empathy, which is still not fully conceptualized but is accepted as a multidimensional structure, have been found (Cuffs et al., 2016). Empathy, in its simplest definition, is the response of one person to the emotions or situation of another (Davis, 1983). Empathy includes cognitive and emotional abilities (Cuff et al., 2016; Hall, & Schwartz, 2019). Cognitive empathy represents the ability to understand another's emotions and adopt their perspective. Emotional empathy is the ability to share another person's feelings. Another type of emotional response, "*empathic concern*" is feeling compassionate feelings for another person who is in distress (Cuff et al., 2016; Davis, 1983; Hall, & Schwartz, 2019).

Although research has historically focused on empathy between humans, humans can also feel empathy for animals (Franklin et al., 2013; Paul, 2000; Tam, 2013; Young et al., 2018). For example, a person may understand the suffering of an injured animal, experience that pain as if he or she were experiencing it himself or herself and feel compassion for that animal. In addition, Tam (2013) introduced the concept of "*empathy with nature*," which suggests that the human capacity for empathy extends to nature and defined it as "*the understanding and sharing of the emotional experience, particularly distress, of the natural world*" (p. 93). According to Tam (2013), people may be encouraged to empathize with nature (induced empathy) or they may have a predisposition to spontaneously empathize with nature as a personality trait (dispositional empathy). However, even if empathy for humans and non-humans (empathy for animals and nature) are related to some extent, there is evidence that these concepts are different psychological phenomena (Gómez-Leal et al., 2021; Sevillano et al., 2017; Paul, 2000; Tam, 2013; Taylor & Signal, 2005). Other studies (Geiger et al., 2017; Kim & Cooke, 2021; Walker & Chapman, 2003) have found evidence to support the idea that people can feel empathy for nature (e.g., ecosystems). Consistent with Tam's (2013) propositions, research has shown that individuals who feel empathy with specific elements of nature (e.g., animals) or with nature as a whole, whether dispositional or induced, are more likely to exhibit nature conservation behavior (Ballarotto et al., 2025; Berenguer et al., 2007; Büscher et al., 2023; Ghasemi & Kyle, 2021; Greving & Kimmerle, 2021; Kim & Cooke, 2021; Maguire et al., 2020; Smith et al., 2024; Swim, & Bloodhart, 2015; Wang et al., 2023; Yin et al., 2021). Empathy may also play an important role in decisions about animal welfare. For example, consumers who feel stronger empathy toward suffering fish report stronger moral obligations to purchase fish with better welfare (Govaerts & Altintzoglou, 2024). Additionally, recent research suggests that empathy plays a critical role in human-wildlife coexistence. For example, empathy for wildlife was associated with greater tolerance toward them (Kansky & Kidd, 2024).

The Dispositional Empathy with Nature (DEN) Scale

Tam (2013) developed DEN scale, which fills an important gap in literature as an original tool to measure individual differences in the tendency to empathize with nature. The DEN scale includes items referring to "*perspective taking*" representing cognitive empathy, and "*empathic concern*"

representing emotional empathy. Since the (DEN) scale is intended to measure empathy for distress in the natural world in general, the scale includes a short instruction that directs participants to think of specific examples of situations in which nature is distressed. Thus, the scale items are associated with a specific context. The items refer to the experiences of animals and plants to represent nature (e.g., "*I imagine how I would feel if I were the suffering animals and plants*"). Participants are asked to indicate whether they agree or disagree with the items on a seven-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). The DEN scale has demonstrated a one-dimensional structure (Tam, 2013), supporting the assumption that the cognitive and emotional dimensions of empathy cannot be empirically separated (Baron-Cohen & Wheelwright, 2004). The one-dimensional structure of the scale has also been confirmed in Spanish (Sevillano et al., 2017) and Italian (Lovati et al., 2025) samples. All or some items of the DEN scale have also been used to measure empathy towards wildlife more specifically. For example, Ghasemi and Kyle (2021) measured empathy towards wildlife in their research. For this purpose, they changed the phrase "animals and plants" in the original items of the DEN scale to "wildlife". Büscher et al. (2023) adapted items from the DEN scale to measure empathy towards "endangered animals in Ecuador". Similarly, Govaerts and Altintzoglou (2024) measured the tendency to empathize with fish using some items from the DEN scale.

This study examined the construct validity and reliability of the DEN scale (Tam, 2013), adapted to measure empathy towards wolves, in a Turkish sample. The gray wolf (Canis lupus) is a native and ecologically important species found across all regions of Turkey (Ambarlı et al., 2016), including Ankara where the study was conducted. Human-wildlife conflicts (Ambarlı, 2019; Thirgood et al., 2005) due to carnivorous nature makes wolves a socially and emotionally salient species for assessing empathy. The study is important in terms of encouraging research on empathy with nature and wild animals in Turkey and testing the cross-cultural validity of the DEN scale.

Methodology

Sample and Data Collection

This research was conducted using a classical printed survey method. The data was collected between March 2024 and May 2024. The sample consisted of 1,140 high school students from different grade levels (Table 1) studying in public schools in 19 different districts of Ankara. The total sample included 618 females, 507 males, and 15 students who did not specify their gender. The students were selected using a convenience sampling method. The ages of the students ranged from 13 to 19 (M=15.72; SD=0.87).

Table 1. *Participants' Grade Levels.*

Grade Level	Total	Percentage
9	228	20
10	516	45
11	347	30
12	49	5
Total	1140	100

Students from different types of high schools participated in the study (Table 2). All schools participating in the study were state schools. The survey was conducted during class hours. Prior to the survey, students were informed about the purpose of the study, and it was emphasized that participation was entirely voluntary.

Table 2. *Types of Schools Participating in the Study.*

School Type	Total	Percentage
Anatolian High School	714	62
Anatolian Imam Hatip High School	94	8
Multi-Program Anatolian High School	42	4
Science High School	235	21
Vocational Technical and Anatolian High School	55	5
Total	1140	100

Adaptation Procedure

The original DEN scale developed by Tam (2013) consisted of 10 items. In our study, all 10 items were retained with only replacing the phrase “animals and plants” with “wolves” to measure dispositional empathy with wolves. Next, the scale items were independently translated into Turkish by three academic experts who completed their doctoral education in the field of science education in the USA and are fluent in both English and Turkish. The translated versions were then reviewed and discussed with one of the translators and the authors of this article, and a consensus was reached on the final version. A short instruction was added to provide context to the items in the scale, including examples of situations in which wolves suffer in nature. Like the original scale, participants were asked what they thought and felt when they heard such news about wolves. Participants were asked to what extent they agreed with the statements on a 5-point Likert scale instead of the original 7-point scale.

Analysis

SPSS 23 statistical program was used in the analysis of the data. The validity of the scale adapted to measure dispositional empathy towards wolves was examined using exploratory factor analysis (EFA). The reliability of the scale was examined by calculating the Cronbach's alpha coefficient and item-total correlations. The independent t-test was used to examine whether there was a difference between male and female students in terms of empathy scores towards wolves.

Results

The KMO coefficient (0.917) and Bartlett's sphericity test (Chi-Square = 6476.269; df=45; p=0.000) results showed that the data were suitable for factor analysis. As a result of the factor analysis, it was determined that all DEN scale items combined under one principal component (Table 3). The single factor explained 57% of the total variance.

Table 3. *Factor Analysis Results.*

Component Matrix[a]	
Item Number	Component 1
8	,821
2	,811
9	,795
3	,794
4	,758
5	,756
1	,731
7	,711
6	,702
10	,689

Cronbach Alpha coefficient of the scale was determined as 0,917. The item total statistics are presented in Table 4. As seen from the Table Cronbach's Alpha value ranges between 0,905 and 0,913 for any deleted item. That shows that all items fit in the scale and there is no need to delete any item.

Table 4. *İtem Total Statistics of DEN Scale.*

Item-Total Statistics				
Item Number	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	32,40	56,657	,662	,911
2	32,52	55,300	,752	,906
3	32,76	54,441	,733	,907
4	32,73	55,028	,694	,909
5	32,34	56,143	,691	,909
6	32,48	56,888	,632	,912
7	32,00	57,390	,639	,912
8	32,80	54,430	,765	,905
9	32,81	54,647	,733	,907
10	31,75	58,334	,616	,913

Skewness (-0.836) and Kurtosis (0.870) values showed that the empathy variable had a normal distribution. As shown in Table 5 independent samples t-test was conducted to compare empathy toward wolves between male and female participants, as measured by the adapted Turkish version of DEN scale.

Table 5. *Independent Sample t-test Results.*

	Levene's Test		**t-test for Equality of Means**				
	F	Sig.	t	df	Sig.	Mean Difference	Std. Error Difference
Equal variances assumed	33,855	,000	8,308	1123	,000	4,063	,489
Equal variances not assumed			8,101	936	,000	4,063	,502

Levene's test for equality of variances was significant ($F = 33.855$, $p < .001$), indicating that the assumption of equal variances was violated. Therefore, results based on the "equal variances assumed" and "equal variances not assumed" rows were interpreted independently.

Table 6 displays group statistics and t-test results for gender differences in empathy toward nature (specifically wolves), using the adapted DEN scale, under the assumption of unequal variances. The analysis revealed that female participants (N = 618, $M = 37.69$, $SD = 7.13$) scored significantly higher on the DEN scale compared to male participants (N= 507, $M = 33.63$, $SD = 9.26$), $t(936) = 8.101$, $p < .001$. These results suggest that females show a significantly greater dispositional empathy toward wolves compared to males. Cohen's $d = 0.49$ was obtained, which is a practically significant difference indicating that the effect size is at a medium level.

Tablo 6. *Group Statistics for not Equal Variances Assumed.*

Gender	**F**	**Sig.**	**Df**	**N**	**t**	**Mean**	**Std. Deviation**
Female	33,855	0.000	936	618	8,101	37,69	7,13
Male				507		33,63	9,26

Discussion

When people's empathy orientations towards wildlife are examined, it is seen that they exhibit positive attitudes and behaviors that protect and support wildlife (Berenguer et al., 2007; Büscher et al., 2023; Ghasemi & Kyle, 2021; Greving & Kimmerle, 2021; Kansky & Kidd, 2024). The DEN scale, which was developed to measure empathy towards nature, has also been seen as a usable, effective and functional tool in assessing empathy felt towards wild animals (Büscher et al., 2023; Ghasemi & Kyle, 2021; Govaerts & Altintzoglou, 2024). However, given that culture shapes both people's empathic responses (Jami et al., 2024) and their relationships with animals (Prato-Previde et al., 2022), such scales can be expected to yield different results in different cultures. This article examines the construct validity and reliability of the DEN scale, adapted to measure empathy towards wolves, in a sample of Turkish adolescents. As a result of the exploratory factor analysis conducted on the modified items to measure empathy with wolves, a single-factor structure was obtained. Therefore, the adapted scale overlapped with the single-factor structure of Tam's (2013) original DEN scale. The analysis conducted to test reliability showed that the scale had high internal consistency. In summary, the adapted DEN scale was found to be a valid and reliable scale in assessing individual differences in the tendency to empathize with wolves in the Turkish sample. These findings confirmed the capacity of the DEN scale to measure empathy towards a specific element of nature, such as a wild animal species. However, the original DEN scale measures empathy with nature. Therefore, our findings cannot be directly generalized to the original DEN scale. The validity and reliability of the original DEN scale within a Turkish sample should be examined independently.

The findings obtained with DEN, which was adapted to Turkish and formatted specifically for wolves, proved that the empathy levels of female participants were significantly higher than those of male participants. This result is consistent with other research showing that females have stronger empathy (Paul, 2000) and moral concern (Herzog, Betchart, and Pittman, 1991; Kellert and Berry, 1987) for animals.

The fact that the sample consisted of adolescents and that the convenience sampling method was used can be expressed as a limitation of this study. Most of the sample consisted of adolescents living in the urban area. Therefore, it may be difficult to generalize our findings to the general population. However, our findings are important in terms of showing that the DEN scale has a robust factorial structure and can measure empathy towards wolves in a Turkish sample.

Conflicts of Interest: No conflict of interest has been declared by the authors.

Funding Details: This study was not funded by any organization.

Ethical Statement: Ethical approval for this research was provided by the Scientific Research and Publication Ethics Committee of Hacettepe University in the field of Social Sciences and Humanities (date and number: 11/08/2023 - 3054376).

Credit Author Statement: Each author made an equal contribution to the study.

Footnote: This research was carried out within the scope of a doctoral study conducted within the Hacettepe University Institute of Educational Sciences. The (DEN) scale used in the study was defined as a functional and meaningful tool in terms of measuring not only general empathy towards nature but also specific empathy felt towards wild animals. It was determined that the scale would make a significant contribution to the use of empathy-based measurement tools in the field of environmental education.

References

Ambarlı, H. (2019). Analysis of wolf–human conflicts: Implications for damage mitigation measures. *European Journal of Wildlife Research*, *65*(6), 81.

Ambarlı, H., Ertürk, A., & Soyumert, A. (2016). Current status, distribution, and conservation of brown bear (Ursidae) andwild canids (gray wolf, golden jackal, and red fox; Canidae) in Turkey. *Turkish Journal of Zoology*, *40*(6), 944-956.

Ballarotto, G., Ghezzi, V., & Velotti, P. (2025). Feeling the nature to foster sustainability: The mediating role of (self) compassion. *Sustainability, 17*(1), 351.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of autism and developmental disorders, 34*, 163-175.

Berenguer, J. (2007). The effect of empathy in proenvironmental attitudes and behaviors. *Environment and Behavior, 39*(2), 269-283.

Brown, K., Adger, W. N., Devine-Wright, P., Anderies, J. M., Barr, S., Bousquet, F., ... & Quinn, T. (2019). Empathy, place and identity interactions for sustainability. *Global Environmental Change, 56*, 11-17.

Büscher, M., Stein, L., Durán, M. E., Cazar, M. E., Hillebrand, P., Schlünder, S., & Fiebelkorn, F. (2023). Ecuadorian children's willingness to protect endangered species–identifying behavioral predictors in a biodiversity, hotspot. *Society and Animals, 1*, 1-24.

Cuff, B. M., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion Review, 8*(2), 144-153.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113.

Dirzo, R., Young, H. S., Galetti, M., Ceballos, G., Isaac, N. J., & Collen, B. (2014). Defaunation in the Anthropocene. *Science, 345*(6195), 401-406.

Franklin Jr, R. G., Nelson, A. J., Baker, M., Beeney, J. E., Vescio, T. K., Lenz-Watson, A., & Adams Jr, R. B. (2013). Neural responses to perceiving suffering in humans and animals. *Social Neuroscience, 8*(3), 217-227.

Geiger, N., Bowman, C. R., Clouthier, T. L., Nelson, A. J., & Adams Jr, R. B. (2017). Observing environmental destruction stimulates neural activation in networks associated with empathic responses. *Social Justice Research, 30*(4), 300-322.

Ghasemi, B., and G.T. Kyle, (2021). Toward moral pathways to motivate wildlife conservation. *Biological Conservation,* 259, 109170. https://doi.org/10.1016/j.biocon.2021.109170

Goleman, D. (1998). *Duygusal Zekâ: Neden IQ'dan daha önemlidir?* (9. Baskı, s. 100-120) İstanbul Varlık Yayınları

Govaerts, F., & Altintzoglou, T. (2024). Consumer intention to buy products containing fish with better welfare: the role of empathy in an extended value–belief–norm model. *British Food Journal, 126*(10), 3684-3698.

Gómez-Leal, R., Costa, A., Megías-Robles, A., Fernández-Berrocal, P., & Faria, L. (2021). Relationship between emotional intelligence and empathy towards humans and animals. *PeerJ, 9*, *e11274*, 14.

Greving, H., & Kimmerle, J. (2021). You poor little thing! The role of compassion for wildlife conservation. *Human Dimensions of Wildlife, 26*(2), *115-131.*

Hall, J. A., & Schwartz, R. (2019). Empathy present and future. *The Journal of Social Psychology, 159*(3), 225-243.

Herzog, H.A., Betchart, N.S., & Pittman, R.B. (1991). Gender, sex role orientation and attitudes toward animals. *Anthrozoös,* 4(3),184-191 https://doi.org/10.2752/089279391787057170

Ives, C. D., Abson, D. J., Von Wehrden, H., Dorninger, C., Klaniecki, K., & Fischer, J. (2018). Reconnecting with nature for sustainability. *Sustainability Science, 13,* 1389-1397.

Jami, P. Y., Walker, D. I., & Mansouri, B. (2024). Interaction of empathy and culture: a review. *Current Psychology, 43*(4), 2965-2980.

Kansky, R., & Kidd, M. (2024). Putting yourself in an animal's shoes-empathy and intangible benefits drive tolerance towards wildlife in Namibian communal conservancies. *Biological Conservation, 293*, 110588.

Kellert, S. R., & Berry, J. K. (1987). Attitudes, knowledge and behaviors toward wildlife as affected by gender. *Wildlife Society Bulletin, 15*(3*)*, 363–371.

Kim, S. C., & Cooke, S. L. (2021). Using the health belief model to explore the impact of environmental empathy on behavioral intentions to protect ocean health. *Environment and Behavior, 53*(8), 811-836.

Lovati, C., Manzi, F., Di Dio, C., Massaro, D., Gilli, G., & Marchetti, A. (2025). Bonding with nature: a validation of the dispositional empathy with nature scale in Italy. *Frontiers in Psychology, 16*, 1388798.

Maguire, P., Kannis-Dymand, L., Mulgrew, K. E., Schaffer, V., & Peake, S. (2020). Empathy and experience: understanding tourists' swim with whale encounters. *Human Dimensions of Wildlife, 25*(2), 105-120.

Paul, E. S. (2000). Empathy with animals and with humans: Are they linked? *Anthrozoös, 13*(4), 194-202.

Prato-Previde, E., Basso Ricci, E., & Colombo, E. S. (2022). The complexity of the human–animal bond: Empathy, attachment and anthropomorphism in human–animal relationships and animal hoarding. *Animals, 12*(20), 2835.

Rifkin, J. (2009). The empathic civilization: The race to global consciousness in a world i crisis. Penguin.

Sevillano, V., Aragonés, J. I., & Schultz, P. W. (2007). Perspective taking, environmental concern, and the moderating role of dispositional empathy. *Environment and Behavior, 39*(5), 685-705.

Sevillano, V., Corraliza, J. A., & Lorenzo, E. (2017). Spanish version of the dispositional empathy with nature scale/versión española de la escala de empatía disposicional hacia la naturaleza. *International Journal of Social Psychology, 32*(3), 624-658.

Schultz, P. W. (2000). New environmental theories: Empathizing with nature: The effects of Perspective taking on concern for environmental issues. *Journal of Social Issues, 56*(3), 391-406.

Smith, P., Mann, J., & Marsh, A. (2024). Empathy for wildlife: The importance of the individual. *Ambio, 53*(9), 1269-1280.

Swim, J. K., & Bloodhart, B. (2015). Portraying the perils to polar bears: The role of empathic and objective perspective-taking toward animals in climate change communication. *Environmental Communication, 9*(4), 446-468.

Tam, K. P. (2013). Dispositional empathy with nature. *Journal of Environmental Psychology, 35*, 92-104.

Taylor, N., & Signal, T. D. (2005). Empathy and attitudes to animals. *Anthrozoös, 18*(1), 18-27.

Thirgood, S., Woodroffe, R., & Rabinowitz, A. (2005). The impact of human–wildlife conflict on human lives and livelihoods. In R. Woodroffe, S. Thirgood, & A. Rabinowitz (Eds.), *People and wildlife: Conflict or coexistence?* (pp. 13–26). Cambridge University Press.

Walker, G. J., & Chapman, R. (2003). Thinking like a park: The effects of sense of place, perspective-taking, and empathy on pro-environmental intentions. *Journal of Park and Recreation Administration, 21*(4), 71–86.

Wang, L., Sheng, G., She, S., & Xu, J. (2023). Impact of empathy with nature on pro-environmental behaviour. *International Journal of Consumer Studies, 47*(2), 652-668.

Yin, C., Ma, H., Gong, Y., Chen, Q., & Zhang, Y. (2021). Environmental CSR and environmental citizenship behavior: The role of employees' environmental passion and empathy. *Journal of Cleaner Production, 320*, 128751.

Young, A., Khalil, K. A., & Wharton, J. (2018). Empathy for animals: A review of the existing literature. Curator. *The Museum Journal, 61*(2), 327-343.

"CheatGPT" or Learning Tool?: Unpacking Student Motivations and Policy Influence in the Age of Generative AI

Gregory Price*
Tokyo University of Science (Japan)

Marc D. Sakellarios**
Toyo University (Japan)

Abstract

The rapid adoption of generative AI in academia forces a critical question. Is the technology a "CheatGPT" or a genuine learning tool? This mixed-methods study investigated the perceptions of Japanese university students in English writing courses (n_1 = 99, n_2 = 96) regarding AI usage by analyzing their motivations, rationale for restraint (ethics), and understanding of institutional policy. Findings indicate an overwhelming majority of participants reported instrumental use of AI but do not feel comfortable doing so. Many reported guilt, and roughly half reported the fear of self-sabotaging their learning skills. Moral ambiguity was common when students were questioned about AI, with participants giving various explanations of what constitutes cheating. The results highlight a strong student demand for clear guidance and suggest that teacher permission is a primary determinant of usage. The study concludes that the lack of universally agreed-upon ethical definitions poses a critical barrier to policy implementation, necessitating immediate, clear, and contextual guidance from instructors and institutions.

Keywords: Generative AI, Academic Integrity, Educational Technology, Plagiarism, Cheating

* ***Gregory Price***, Department of Science and Engineering, Tokyo University of Science, Japan, ORCID ID: https://orcid.org/0000-0002-3635-5060 e-mail: grendel.t@rs.tus.ac.jp

** ***Dr. Marc D. Sakellarios*****,** Faculty of Life Sciences, Department of Life Sciences, Toyo University, Japan, ORCID ID: https://orcid.org/0009-0007-1254-4020 Email: sakellarios@toyo.jp

Correspondence**:** grendel.t@rs.tus.ac.jp

Introduction

The researchers, prior to the conception of this paper, met and informally discussed issues they have had in their English writing courses in Japan. One brought up an experience of a student turning in a paper with a reference section in which several sources listed indicated the wrong authors. When the teacher questioned the student about this error, the student replied, "This is what ChatGPT said." This apparent lack of comprehension regarding the origin, verification, and authorship of the content sparked the curiosity that led to this study.

The rapid adoption of generative AI (GenAI) challenges not only traditional concepts of academic integrity but also the fundamental nature of student-teacher dynamics. Academic integrity has long been a foundational concern, but the current accessibility of tools like ChatGPT requires a fundamental reassessment of what constitutes original work and cheating. This challenge is particularly acute in international English as a Foreign Language (EFL) settings where, apart from evolving policies, students already navigate a complex linguistic and cultural landscape.

Existing literature confirms the high prevalence of AI use but often fails to integrate internal dilemmas (guilt, self-sabotage) and external factors (teacher policy, institutional influence) that govern student usage decisions. To address this gap, the present study employs a mixed-methods approach to investigate three central themes influencing student behavior in a Japanese university setting: motivation and efficacy, the rationale for restraint, and teacher/institutional influence. These emergent themes are explored through the organizing questions: "Why?", "Why Not?", and "Where's the Line?". Following this introduction, the paper summarizes relevant literature, details the mixed-methods survey, presents the quantitative and qualitative findings, and discusses the implications for policy and pedagogy in the age of AI.

Literature Review

Academic misconduct has long been a concern in education. Carlson (2021) points out that historically, a considerable number of students have engaged in acts of dishonesty, including copying, failure to properly cite borrowed work, and patchwork plagiarism, which is defined as "taking text portions from several different sources, combining them, and presenting the resulting text as one's own work" (Šupak Smolčić & Bilić-Zulle, 2013, p. 17). More recent studies suggest that basic forms of academic misconduct, such as copying, submitting someone else's work, and using "cheat sheets" during tests, have remained essentially the same even as technology has advanced (Carlson, 2021; Gettings & Hughes, 2014; Rinnert & Kobayashi, 2005; Wheeler, 2009). However, emerging technologies such as GenAI introduce a different kind of challenge, one that is less about deception and more about redefining authorship and originality. Cavazos et al. (2024), citing Firat's (2023) study examining the opinions of both educators and students on ChatGPT, suggested that "its mere presence is impactful enough to restructure traditional roles in educational systems" (p. 358).

Evolving Perceptions of Cheating in the Digital Age

Generative AI and Student Ethics

Digital tools can serve as legitimate educational support or convenient ways for creating shortcuts to educational goals that may undercut the benefits of learning and problem-solving (Mok, 2024). Conventional understandings of "cheating," traditionally associated with negative intent, are increasingly complicated by the emergence of AI-assisted academic work. Technology and cultural norms change through generations, resulting in varying perceptions of ethics regarding the use of certain tools in academics; especially when students have had different learning experiences than their teachers (Dawson et al., 2024).

Remote Learning and the COVID-19 Context

When many face-to-face classes were cancelled during the COVID-19 pandemic, online courses became critical for providing students with ongoing education. This resulted in a new normal for the use of digital tools in education (Costa et al., 2021; Koh & Daniel, 2022; Li, 2022). Extended periods of remote learning, where direct supervision was limited, made reliance on digital tools commonplace. This increased students' comfort with online work, establishing a foundation for contemporary attitudes toward the broader use of GenAI for academic purposes (Costa et al., 2021; Graf, 2024; Li, 2022). The prevalence of online education that a generation of students faced during the COVID pandemic is important to consider when developing an understanding of their current perspectives on the proper utilization of AI for academic purposes.

Adoption of GenAI Among Japanese University Students

In the first year of ChatGPT's release, surveys indicated that nearly half of Japanese university students had used GenAI with schoolwork for help summarizing texts, writing or correcting assignments, and assistance with homework (Data Scientist Society, 2025; Mok, 2024; Ohmori et al., 2023). By early 2024, Japan was one of the top five countries using ChatGPT, with an increasing number of those users being of university age (National Federation of University Co-operative Associations, 2024; Price, 2024). In a recent study by Pew Research Center (2025), 77% of young people in Japan aged 18–34 reported a high level of awareness about GenAI, the highest rate among the 25 countries surveyed.

Teacher and Institutional Influences

EFL teachers in Japan (as well as other countries) generally come from diverse cultural backgrounds and are likely to have differing views on the ethical and practical uses of AI in the classroom (Neff et al., 2024b). Students tend to look to teachers for cues on the acceptance of AI usage. Teachers' attitudes towards technology, whether implied or explicit, often shape the students' understanding of AI acceptance on a class-by-class basis. (Luo, 2025; Mah et al., 2024; Neff, et al., 2024b).

Over the last few years, GenAI has swiftly become an undeniable element in educational environments. AI-based tools have become increasingly embedded in everyday digital environments, such as Microsoft 365, Grammarly, and Google Docs, making bans almost impossible to enforce, which underscores the need for clear discussion between educators and students regarding ethical use (Mah et al., 2024; Neff et al., 2024a). For teachers in Japan to meaningfully develop consolidated educational approaches that are both fair and pedagogically effective, it is imperative to understand contemporary student perceptions of proper and improper uses of AI, particularly given the widespread adoption of GenAI among Japanese university students and the new digital norms shaped by recent educational experiences.

Motivation, Ethics, and Policy in Academic Integrity

The following discussion revisits the three guiding themes previously introduced: "Why?", "Why Not?", and "Where's the Line?". We will explore existing literature focusing on how student perspectives on GenAI use intersect with motivation, ethics, and policy.

Why?

There are several theories as to why a student would cheat on academic work. The basic underlying motive might be the fear of academic failure. This is of particular relevance to our study because, according to Diekhoff et al. (1999), Japanese students have a higher degree of expectations than some other cultures to succeed in school, and as such, "experience more pressure to cheat" (p. 351). Wheeler (2014) points out that in the context of EFL classes in Japan, plagiarism can be

motivated by "fear of failure due to struggles with a foreign language" (p. 44). The motivation to cheat can be multifaceted. Yu et al. (2017) found that students with little time for schoolwork due to over-involvement with extracurricular activities tend to engage in "academic misconduct." This should be considered in the case of Japanese students, as a survey by the National Federation of University Co-operative Associations (2025) found that in 2024, 63.8% of Japanese university students took part in club activities, with 80% of first-year students reporting involvement.

Graf (2024) mentions several reasons that cheating through various means, including plagiarism or having a GenAI program write an assignment, might be seen as a useful endeavor for students. One reason is laziness or procrastination, which can create a high-pressure situation. Another motivation being finances, based on the cost/benefit choice of failing a class versus passing and not having to spend money to repeat it. Other students in Graf's (2024) study refer to frustration with working so hard to do well, when it is apparent that other students are cheating and getting similar grades with less effort. Several studies cited by Cavazos et al. (2024), including McCabe et al. (2001) and Rettinger and Kramer (2009), concur that if cheating is normalized in a student's educational environment, then a non-cheater feels disadvantaged.

In a study of students at Rikkyo University in Tokyo, Mok (2024) relays that the most prevalent reasons for using GenAI are for saving time, double-checking facts, as well as to "create content and even write graded assignments" (p.32). If students are going to cheat, GenAI is an optimized means of doing so. Students' motivations to cheat can stem from many factors, including pressure to succeed, limited time, and frustration with an unfair system. Ultimately, such behavior often reflects an attempt to manage overwhelming expectations or to find a more efficient path to success rather than a deliberate intent to deceive (Mok, 2024). These motivations, combined with technological familiarity, directly influence students' reasoning as to why they would use AI assistance with their work.

Why Not?

As GenAI becomes more commonplace, students face novel challenges to the question of ethics in regard to authorship and originality, as well as the possibility of self-sabotaging their educational goals. Students often rationalize AI use pragmatically, citing heavy workloads, using it to help them understand lessons, or simply adapting to an increasingly technological learning environment (Graf, 2024; Dawson et al., 2024). However, several studies and institutional reports highlight a growing awareness among both students and educators that overreliance on GenAI may diminish critical thinking and writing development. Mok (2024) found that Japanese university students themselves recognized that "AI is doing most of the thinking," negatively affecting their cognitive engagement.

Similarly, university guidelines across Japan, such as those issued by the University of Tokyo and Okayama University, warn that the convenience of AI-generated content risks depriving students of meaningful learning experiences, reinforcing the perception that excessive dependence on AI can hinder intellectual growth (Abir & Zhou, 2025).

Research shows that students have some ability to judge what constitutes ethical AI usage (Graf, 2024; Mok, 2024; Neff et al., 2024b). Their ethical judgments on AI use are often tied to the perceived fairness, individual effort, and resulting feelings of moral wrongdoing (guilt) concerning the work's authorship and adherence to teacher expectations concerning the extent to which they should (or should not) use AI assistance (Cavazos et al., 2024; Golding et al., 2025; Welding, 2023). If the teachers aren't clear about their expectations regarding AI use, the students can treat the ambiguity as indirect permission to use it in ways that may cross ethical boundaries (Graf, 2024). Because many students were educated in online environments during the COVID-19 pandemic, practices such as copying and pasting material for assignments may not carry the same sense of wrongdoing that traditional plagiarism once did (Dawson et al., 2024; Graf, 2024).

Most students represented in studies appear ready and willing to use AI responsibly when clear academic standards are provided, understanding that using AI-generated work without attribution is improper (Cavazos et al., 2024; Chan, 2024). Certain utilizations, such as citation generators, generating paraphrases or summaries, and generally integrating AI output into their own writing, are seen by students as gray areas between legitimate support and possible misconduct (Graf, 2024; Mok, 2024). These findings suggest that students are weighing convenience against academic norms in the context of the assignment and teacher expectations (Mok, 2024; Neff et al., 2024b; Welding, 2023).

Historically, plagiarism among Asian students has been misattributed to cultural norms, but recent research shows that Japanese students typically understand what constitutes plagiarism, and that the issue lies instead with a lack of training in citation and referencing norms in foreign-language academic writing (Liu, 2005; Pecorari, 2003; Pennycook, 1996; Sowden, 2005). Japanese schooling emphasizes accuracy and reproduction, with a strong focus on preparing for national entrance examinations that rarely require citation in essays. Consequently, explicit instruction in proper referencing, paraphrasing, and summarizing is often limited before university, leaving many students with only basic experience in integrating sources (Rinnert & Kobayashi, 2005; Sadoshima, 2014; Teeter, 2014). This context helps explain why Japanese students may struggle with citation methods and academic writing standards, especially when using English, even though they seem to exhibit ethical awareness regarding plagiarism.

Adding to the issues of EFL students working with academic writing in English, the incorporation of GenAI complicates ethical scholarship by blurring the line between assistance, authorship, and misconduct (Graf, 2024; Dawson et al., 2024). This technological shift has created practical challenges for educators, as recent research shows that AI-detection tools often produce inconsistent results, leaving human judgment as the more reliable means of assessment (Price & Sakellarios, 2025). In this context, EFL instructors vary in their approach. Some treat situations that arise, such as inadequate citation, as an ethical violation, while others use it as an opportunity to provide targeted guidance on academic writing standards (Rinnert & Kobayashi, 2005; Teeter, 2014). These variations in instructor approach highlight the broader role that classroom practices and institutional policies play in shaping how students understand and navigate the boundaries of acceptable GenAI use.

Where's the line?

Institutional and instructor policies strongly shape how students interpret academic integrity in the age of AI. In Japan, policy development around GenAI has so far followed a more decentralized and pedagogically oriented path. Rather than imposing uniform prohibitions, Japanese universities tend to emphasize process-oriented learning and student responsibility over punitive enforcement (Abir & Zhou, 2025; Mack & Byanjankar, 2025). The Ministry of Education, Culture, Sports, Science and Technology (MEXT) issued general guidance in 2023, allowing universities to determine their own approaches based on educational context (MEXT, 2024). As a result, roughly three-quarters of leading institutions have implemented GenAI policies, many of which delegate interpretive authority to individual instructors (Abir & Zhou, 2025; Saito, 2024).

Universities such as the University of Tokyo, University of Tsukuba, Yokohama National University, and Keio University clearly connect their GenAI guidelines to learning objectives, cautioning that over-dependence on AI may undermine students' critical thinking and responsibility of authorship (Keio University, 2023; University of Tokyo, 2023; University of Tsukuba, 2023; Yokohama National University, 2023). This autonomy-driven model perhaps reflects a broader Japanese cultural preference for contextual ethics, in which instructors are entrusted to balance an openness for technological innovation with academic integrity within their own classrooms.

Even with clearly articulated institutional policies, the real test of academic integrity occurs in the classroom, where instructor attitudes and practices determine how such policies are understood in practice. Research shows that students often take cues from their professors' behavior and

expectations. When instructors emphasize the educational purposes behind integrity policies, students are less likely to view misconduct as acceptable (Graf, 2024; Hutton, 2006; Vandehey et al., 2007). Graf (2024) refers to this as the classroom "vibe," a combination of clear rules, instructor involvement, seriousness, and apparent dedication, which signals whether shortcuts or unethical practices are tolerated. As Neff et al. (2024b) observed in Japanese EFL settings, teachers' cultural and ethical frameworks strongly influence their interpretations of AI-assisted writing. Consequently, student perceptions of "cheating" often depend as much on instructor guidance, the classroom "vibe," and contextual expectations as on institutional policy. This interplay between institutional autonomy and classroom-level discretion highlights the complexity of developing coherent academic integrity practices in Japan's rapidly evolving educational environment.

Methods

To better understand student motivations to use AI, their reasons to abstain from AI use, and student understanding of teacher and institutional policies, two questionnaires were utilized to elicit a range of qualitative and quantitative data from Japanese university EFL student participants, who were either studying academic writing or taking courses that involved writing at a scholarly level. This methodology, which intentionally combines numerical data with textual responses, constitutes a mixed-methods approach.

The first questionnaire had 20 items regarding AI student use and incorporated a 10-point Likert scale. The items were categorized based on three themes: 1. Why? (motivation to use AI), 2. Why not? (reasons to abstain), and 3. Where's the line? (students' understandings of teacher and institutional policies). The researchers categorized questionnaire items based on their initial research question: Is the technology a "CheatGPT" or a genuine learning tool? Essentially, the researchers strived to understand the extent and limiting boundaries self-applied by students utilizing GenAI in an academic setting.

The questions were then randomized and copied into a Google Forms site. Japanese translations were written under the English questions to mitigate misunderstandings due to language barriers. A QR code linking student participants to the questionnaire was distributed to 612 university students across 6 Japanese universities.

Despite conventional wisdom suggesting that a 10-point scale creates high response effort and perhaps higher non-response rates, Coelho and Esteves (2007) found, perhaps counterintuitively, higher non-response rates in five-point scales. Additionally, Pearse (2011) recommends that researchers offer respondents finer granularity when designing questionnaires. The researchers of the present study felt that a 10-point scale would offer respondents finer granularity when responding to questions regarding nuanced concepts of academic honesty.

Additionally, a second open-ended questionnaire was given to triangulate data from the Likert-type survey and supply deeper qualitative insights to either support or counter quantitative findings. The same participants were given both questionnaires at the same time. The second questionnaire had two open-ended items with Japanese translations under them. The participants were advised to answer either in English or Japanese. These questions were as follows:

1.) In your opinion, where is the line between getting help and cheating when using digital tools like ChatGPT or online translators for academic work?

2.) Do you believe AI tools will improve or weaken student learning over time? Support your opinion with examples from your experience.

A separate Google Form was used for this second questionnaire, and student participants were given corresponding QR codes and asked to complete both questionnaires in the same time period.

Results

In this section, we will discuss the results of the 20-item 10-point Likert questionnaire. This section will be divided into three parts: 1. Why? (Prevalence and Motivation), 2. Why Not? (Reasons to Abstain from AI Usage), and 3. Where's the Line? (Teacher and Institutional Policies). The response rate to the first questionnaire was 16.1% ($n_1 = 99$). Qualitative data from the second questionnaire will also be used to elucidate data garnered from the closed-ended first questionnaire.

The response rate to the second open-ended questionnaire was 15.6% ($n_2 = 96$). Essentially, three participants who completed the Likert-style questionnaire did not complete the second questionnaire. This section will be followed by a brief discussion unpacking the researchers' views of the results.

Why?: Prevalence and Motivation; Perceived Efficacy and the Future of Use

When participants were presented with the question, "Have you ever used ChatGPT or a similar AI tool to help with an English writing assignment?", it was interesting to discover that 92% of respondents had used it on some level. The main reasons for using AI were "improving grammar" (46.9%) and "getting ideas" (40.6%), with "saving time" less common (12.5%). This aligns with Mok (2024), who found that students used AI tools instrumentally rather than deceptively. As Mok (2024) states, "When used for learning purposes, students used it mostly to deepen their understanding of subject matter" (p. 32). It is still alarming that 12% of participants use AI simply to save time. This may imply that learning is not a primary objective of their AI use. This time-saving motivation may connect with the Yu et al. (2017) finding that Japanese students' time was compromised due to heavy engagement in extracurricular school activities.

Instrumental use of AI as a learning tool was supported by the data in the open-ended questionnaire. As one participant states, "I think that using it to look up English expressions or the meanings of terms can improve time efficiency and lead to improved learning" (translated from Japanese). Another participant adds, "If you think carefully yourself and only use AI tools when you absolutely have to, I think it will lead to improved learning" (translated from Japanese). Conversely, there was a response that indicated selective academic dishonesty rather than benign motivations. As the participant suggests:

> I often leave everything to AI for classes I'm not interested in but must take, or for reports that are completely useless to me. However, for classes that I think are beneficial to me (such as English and French), I only use AI tools for things I don't understand or questions I have, and I use my own knowledge for the rest (translated from Japanese).

Another participant adds, "In my own case, using AI has saved me time, so I've been able to devote that extra time to other self-study" (translated from Japanese). The same participant also warns, "[…] since AI makes the process more efficient and reduces the overall study time, students may feel less sense of accomplishment, which could negatively affect their motivation to learn" (translated from Japanese). One participant expresses the time benefits of AI by stating, "Translation can be done immediately without the need to use other electronic devices or dictionaries on your smartphone, making it time efficient" (translated from Japanese).

When presented with the question "Do you feel AI tools make learning English easier or harder?" the median score was 7, indicating a strong agreement with the premise that AI makes learning easier. In fact, 72.8% of participants answered with 6 or above, indicating a general belief in the efficacy of AI tools in facilitating learning. This quantitative finding was also reflected in the open-ended survey data. For instance, one participant explained the utility, stating, "It gives you many ideas when you input a command" (translated from Japanese). However, the same participant also highlighted a key drawback: "[…] sometimes you mistakenly think those ideas [ones garnered from AI] are your own," suggesting an awareness of the challenge of accurately crediting intellectual

property. Another participant similarly pointed to the need for self-regulation when using these tools: "Some people have self-control and can use it appropriately, while others are easily swayed by the easier option of not thinking for themselves at all" (translated from Japanese). This participant went on to suggest that long-term exposure to AI may help students understand appropriate boundaries, writing, "I also think that by being exposed to AI from an early age, people can learn how to use it effectively" (translated from Japanese).

Additionally, when presented with the question "In the future, how often do you think students will use AI for academic writing assignments?", 89.8% answered with a score of 6 or above, indicating general agreement. This suggests that the participants not only feel AI is useful but expect its usage to be increasingly prevalent in academia in the future.

Additionally, there was a sense among many participants that students in the future would become overly dependent on AI. As one participant writes, "I worry that we might become overly reliant on AI tools in the future" (translated from Japanese). Another participant cautions, "[…] over-reliance [on AI] will eliminate opportunities to develop independent thinking" (translated from Japanese).

Why Not?: Reasons to Abstain; The Role of Moral Ambiguity, Guilt, and Self-Sabotage

The participants questioned do seem to possess a moral compass about the ethics of cheating, with 90.1% agreeing to the statement "Cheating is always wrong" (quantified as answers 6 or above). However, the line regarding what qualifies as cheating becomes blurred with AI. When asked what they considered to be cheating, only 62.6% of respondents agreed that "Using ChatGPT to write an essay" constituted cheating. This suggests some moral ambiguity regarding using AI to write essays. Indeed, this might suggest that the remaining 37.4% of respondents do not see AI-generated writing as cheating. It is worth noting that the question was stated as "Using AI to *write* an essay", rather than using it to help with understanding or as a tool for inspiration. Neff (2024b) suggests there might be a dichotomous understanding of the ethical use of AI harbored by teachers versus students. As Neff (2024b) states, "students considered using AI to complete an assignment as acceptable, but their EFL teachers thought this was unfair and unethical because they expected students to complete homework without using AI" (p. 61).

This ethical uncertainty or potential confusion was reflected in answers on the open-ended survey. Some participants seemed to feel that using it from the onset was problematic, but introducing it later was acceptable. As one participant expresses, "I believe that using them [AI tools] from the start, without thinking for myself, is cheating" (translated from Japanese). As another participant echoes, "Using it from the start is cheating" (translated from Japanese). Another participant posits, "Using it [AI], despite a notice that it is not allowed, is cheating" (translated from Japanese), suggesting it is acceptable until told otherwise. These answers suggest unclear rationales or guidelines for what constitutes cheating.

A question regarding guilt, worded as "How guilty would you feel if you submitted an AI-written essay as your own work", may offer a glimpse as to why some students refrain from AI usage. Most, or 75.5%, of respondents report they would feel some level of guilt (with scores of 6 or above). There was some connection, in the minds of participants, to using AI and academic dishonestly; but the point at which it became dishonest varied between participants. As one put it, "I think the boundary lies in whether you copy and paste directly or not. Copying and using it directly is cheating" (translated from Japanese). Another participant writes, "I think it's cheating to use content created by AI for an assignment without coming up with the ideas yourself*"* (translated from Japanese), again indicating that ideas can be the intellectual property of AI or those AI is referencing. As stated above, participants sometimes felt it was only cheating to use AI at the beginning, as reflected in this comment, "Having AI think from the ground up is cheating" (translated from Japanese). This overall finding demonstrates that despite high moral standards, the exact line between a helpful learning tool

and cheating remains blurred. This ambiguity suggests that students lack the clear, universal ethical framework necessary to self-regulate usage, underscoring the necessity of external guidance.

The question "Do you think using AI tools prevents you from improving your own writing skills?" adds nuance to the question of guilt. 50.6% suggested it could impede their writing progress. 35.4% agreed somewhat strongly (indicated by answers 8 or above). This suggests perhaps guilt may not be simply rule-based but also mixed with feelings of self-sabotage. In other words, participants may feel they are depriving themselves of an opportunity to better their academic skills, effectively hindering their growth as students.

This was reflected in answers to the open-ended survey. As one participant writes, "In the long term, I think it hinders learning improvement. Once you use AI, you realize how easy it is, and you become reliant on it." (translated from Japanese). This participant goes on to say, "As a result [of using AI], you have fewer opportunities to think for yourself, which I think hinders learning improvement" (translated from Japanese). Another participant adds to this idea by stating, "Although it's very convenient, I think it prevents you from developing your own thinking skills" (translated from Japanese). Yet another echoes this warning, "Because we are using AI, we are not thinking for ourselves." (translated from Japanese). There were indeed many participants who reported concerns that AI would hinder human thinking through over-reliance. Collectively, these answers strongly suggest that the perceived detrimental effect of AI on personal growth serves as a powerful motivator for student restraint.

Where's the Line?: The Demand for Teacher Guidance and the Influence of Institutional Permission

When presented with the question "How comfortable are you discussing AI tools like ChatGPT with your teacher?", 67.8% indicated agreement with scores 6 and above. This indicates, on the part of the student, a willingness to discuss usage and perhaps have parameters set regarding proper use. A surprisingly large number of participants, 25 (25.3%), answered with a "10", perhaps indicating a high interest in having these discussions. This is echoed in one participant's response, who states, "[…]appropriate guidance from teachers to students is necessary when using it [AI] in education*"* (translated from Japanese).

The participants express a desire to receive explicit instruction, with 76.7% agreeing to the statement "Would you like to receive instruction on how to use AI tools ethically for academic purposes?" One participant expresses a concern for proper guidance by stating, "I don't think there's a definitive rule [regarding AI usage] that should always be followed*"* (translated from Japanese), supporting the idea that perhaps moral ambiguity exists when determining AI cheating.

Finally, when participants were presented with the question "If your teacher allows you to use AI tools, would you be more likely to use them?", 59.5% indicated agreement (answers 6 or above). One participant expresses a necessity to use AI based on human limitations, stating, "As a human being, it's impossible to know everything, and I feel it's impossible to read all the books and literature in a short period of time" (translated from Japanese). This participant also warns, "However, there's a possibility that the information isn't accurate, and the information is often incorrect." This indicates that while participants are willing to follow teacher permission, their use remains tempered by their own critical assessment of the tool's inherent limitations regarding factual accuracy.

It was unclear where the exact line between a learning tool and cheating should be drawn. This lack of a universally accepted understanding of what constitutes cheating may make it difficult to implement guidelines institutionally. More research is needed regarding universally agreed-upon definitions of AI cheating.

Discussion

Our findings suggest that students are already widely using AI tools in the classroom. An overwhelming majority reported using or having used AI for assignments. Predominantly, usage was instrumental in purpose, aimed at deepening understanding or generating ideas rather than for outright deception. It is important to note that generating ideas can still be a moral gray area, as ideas can be considered intellectual property. It is not, to the authors' awareness, fully understood how ideas can be "owned" or credited to AI. There was some understanding, on the part of participants, that direct quotation, using AI from the onset, or stealing ideas from AI constituted cheating. This new horizon of AI-generated writing seems to provide more questions than answers regarding idea ownership in the realm of machine intelligence, and more research on this topic is needed.

A clear finding from the data presented in this paper is the participants' demand for guidance in how to properly use AI tools. We find there is both a lack of explicit instruction and a clear understanding on the part of students as to where the line of cheating exists. Providing this guidance could help alleviate feelings of guilt among students who wish to use AI ethically. Since the results suggest most participants are already using AI instrumentally rather than deceptively, it is arguably in everyone's best interest to create clear guidelines for proper usage. If institutions and teachers provided these guidelines, students could use AI somewhat guilt-free to enrich their learning experience. The participants clearly articulated a caveat: the risk of over-reliance on AI and the potential for lazy thinking or hindering critical thinking skills. The ideal goal of AI usage, we feel, should be the enrichment of learning rather than cheating or academic trickery.

Conclusion

The findings of this study provide a comprehensive picture of Japanese university EFL students' complex relationship with GenAI, confirming that usage is widespread and driven by instrumental motivations (such as checking grammar and generating ideas) rather than purely deceptive intent. The central tension revealed across the data lies in the ethical ambiguity students face. While most participants report feeling guilt for submitting AI-written work, a significant portion do not define using AI to generate an essay as cheating. This complexity is reinforced by a reported fear of self-sabotage, with participants expressing concern that using AI impedes the improvement of their own writing skills. This fear acts as a powerful internal restraint, leading some students to restrict usage to avoid diminishing their own cognitive and linguistic development.

Implications for Practice and Policy

The lack of universally agreed-upon ethical boundaries is one of the most significant takeaways from this research. The high student demand for explicit instruction and the strong influence of teacher permission on usage strongly suggest that decentralized institutional policies are insufficient without clear, classroom-level guidance. The instructor, therefore, becomes the primary determinant of ethical behavior in practice. Teachers must move beyond simply policing AI to defining ethical, pedagogical utilization. We recommend institutions prioritize training teachers to provide pedagogical guidance on authorship, originality, and citation methods, thereby mitigating student guilt and reducing the risk of self-sabotage.

Additionally, each university should require teachers to detail explicit policy regarding the use of AI in the syllabus contextually for each course. This removes the current widespread moral ambiguity. It is our recommendation that the focus shift from punitive enforcement to encouraging critical engagement with the tool. By setting clear parameters, educators can mitigate student guilt and facilitate responsible academic growth.

Limitations and Future Research

This study's generalizability is limited by its specific focus on EFL students in Japanese universities and its necessary reliance on self-reported data. While the findings provide rich, contextual insights into the motivations and ethical conflicts of this student population, they may not translate directly to different academic or cultural environments. Furthermore, the findings are cross-sectional, capturing perceptions at a single point in time, which means they cannot account for how policies or technological changes might alter student behavior over the course of an academic year. Future research should expand upon these findings in several ways. Comparative studies across different Asian or Western university systems could validate the moral ambiguity and self-sabotage concerns observed here. Additionally, longitudinal studies are needed to explore the long-term effects of GenAI use, specifically measuring the actual impact of over-reliance on students' linguistic and critical thinking skills. Finally, a corresponding qualitative study examining teacher perceptions of cheating, policy enforcement, and guidance effectiveness would complete the picture of the student-teacher dynamic. Ultimately, the development of effective, non-punitive guidelines hinges on further research that establishes a clear, working definition of "idea ownership" in the era of machine intelligence.

Conflicts of Interest: No conflict of interest has been declared by the authors.

Funding Details: This study was not funded by any organization.

Ethical Statement: The study was conducted following standard ethical research practices. All survey responses were collected anonymously.

Credit Author Statement: Each author made an equal contribution to the study.

References

Abir, M. G. H., & Zhou, K. Z. (2025, August 1). Examining generative AI policies in Japanese universities: A qualitative perspective. Available at SSRN. https://doi.org/10.2139/ssrn.5374222

Carlson, G. D. (2021). Are You Honest? A Study of Self-Reported Academic Misconduct Among Japanese University English Language Learners. *大手前大学IIE ジャーナル [Otemae University IIE Journal]*, *7*, 7-19. https://otemae.repo.nii.ac.jp/records/2206

Cavazos, J. T., Hauck, K. A., Baskin, H. M., & Bain, C. M. (2024). ChatGPT goes to college: Exploring student perspectives on artificial intelligence in the classroom. *Teaching of Psychology.* https://doi.org/10.1177/00986283241268829

Chan, C. K. Y. (2024). Students' perceptions of 'AI-giarism': Investigating changes in understandings of academic misconduct. *Education and Information Technologies*, 29, 1–22. https://doi.org/10.1007/s10639-024-13151-7

Coelho, P. S., & Esteves, S. P. (2007). The choice between a five-point and a ten-point scale in the framework of customer satisfaction measurement. *International Journal of Market Research*, *49*(3), 313-339. https://doi.org/10.1177/147078530704900305

Data Scientist Society. (2025, April 21). Data of Data Scientist シリーズ vol. 65「47%—大学生の生成AI利用率」 [Data of Data Scientist series vol. 65: 47%—University students' generative AI usage rate]. *DSS Journal*. Retrieved September 7, 2025, from https://www.datascientist.or.jp/dssjournal/dssjournal-4013/

Dawson, P., Bearman, M., Dollinger, M., & Boud, D. (2024). Validity matters more than cheating. *Assessment & Evaluation in Higher Education*, 49(7), 1005-1016. https://doi.org/10.1080/02602938.2024.2386662

Diekhoff, G. M., LaBeff, E. E., Shinohara, K., & Yasukawa, H. (1999). College cheating in Japan and the United States. *Research in Higher Education*, 40(3), 343-353. https://doi.org/10.1023/A:1018703217828

Firat, M. (2023). What ChatGPT means for universities: Perceptions of scholars and students. *Journal of Applied Learning & Teaching*, 6(1). https://doi.org/10.37074/jalt.2023.6.1.22

Golding, J. M., Lippert, A., Neuschatz, J. S., Salomon, I., & Burke, K. (2025). Generative AI and college students: Use and perceptions. *Teaching of Psychology*, 52(3), 369-380. https://doi.org/10.1177/00986283241280350

Graf, A. E. (2024). "You Do You": Digital-Native College Students' Perceptions of Cheating and Plagiarism [Doctoral dissertation, The George Washington University]. ProQuest Dissertations Publishing. https://www.proquest.com/dissertations-theses/you-do-digital-native-college-students/docview/2895144273/se-2

Hutton, P. A. (2006). Understanding student cheating and what educators can do about it. *College Teaching*, 54(1), 171–176. https://doi.org/10.3200/CTCH.54.1.171-176

Keio University. (2023, July 2). Guideline for generative AI usage at Keio University. https://www.itc.keio.ac.jp/en/software_ai_guideline.html

Liu, D. (2005). Plagiarism in ESOL students: Is cultural conditioning truly the major culprit? *ELT Journal*, 59(3), 234–241. https://doi.org/10.1093/elt/cci043

Luo, J. (2025). How does GenAI affect trust in teacher-student relationships? Insights from students' assessment experiences. *Teaching in Higher Education*, 30(4), 991-1006. https://doi.org/10.1080/13562517.2024.2341005

Mack, L., & Byanjankar, A. (2025). Understanding policy responses to generative AI in private higher education: Insights from Japan and the United States. *Journal of Academic Ethics*, 1, 1–21. https://doi.org/10.1007/s10805-025-09655-7

Mah, C., Walker, H., Phalen, L., Levine, S., Beck, S. W., & Pittman, J. (2024). Beyond CheatBots: Examining tensions in teachers' and students' perceptions of cheating and learning with ChatGPT. *Education Sciences,* 14(5), 500. https://doi.org/10.3390/educsci14050500

McCabe, D. L., Trevino, L. K., & Butterfield, K. D. (2001). Cheating in academic institutions: A decade of research. Ethics & Behavior, 11(3), 219–232. https://doi.org/10.1207/S15327019EB1103_2

Ministry of Education, Culture, Sports, Science and Technology. (2024). *2024 White Paper on Science, Technology, and Innovation*. https://www.mext.go.jp/en/content/20240611-ope_dev03-000036407-1.pdf

Mok, J. (2024). Students' use of generative artificial intelligence (ChatGPT) at Rikkyo University. 外国語教育研究ジャーナル *[Journal of Foreign Language Education Research], 5,* 21–39. https://doi.org/10.14992/0002001290

National Federation of University Co-operative Associations. (2024, April 8). Half of Japanese university students have used generative AI. *Nippon.com*. https://www.nippon.com/en/japan-data/h01946/

National Federation of University Co-operative Associations. (2025). *Summary report of the 60th fact-finding survey on student life*. https://www.univcoop.or.jp/press/life/report.html

Neff, J., Arciaga, K., & Burri, M. (2024a). EFL students' and teachers' perceptions of the ethical uses of AI tools. *Technology in Language Teaching & Learning*, 6(3), 1714-1714. https://doi.org/10.29140/tltl.v6n3.1714

Neff, J., Arciaga, K., & Burri, M. (2024b). The impact of EFL teachers' demographic backgrounds on the perceptions of ethical AI use. *PanSIG Journal*, 10(1), 60–66. https://doi.org/10.37546/JALTPanSIGJ10.1-8

Ohmori, F., Saitō, J., Matsuba, R., & Kita, T. (2023, June 8). 大学生のChatGPT利用状況と能力形成への影響に関する調査結果（速報） [Survey results on university students' ChatGPT use and impact on skill formation (Preliminary report)]. *DBER Center*. https://dber.jp/chatgptsurvey/

Pearse, N. (2011). Deciding on the scale granularity of response categories of Likert type scales: the case of a 21-point scale. *Electronic Journal of Business Research Methods*, *9*(2), 159-171. https://academic-publishing.org/index.php/ejbrm/article/view/1278

Pecorari, D. (2003). Good and original: Plagiarism and patchwriting in academic second-language writing. Journal of Second Language Writing, 12(4), 317-345. https://doi.org/10.1016/j.jslw.2003.08.004

Pennycook, A. (1996). Borrowing others' words: Text, ownership, memory, and plagiarism. *TESOL Quarterly*, 30(2), 201–230. https://doi.org/10.2307/3588141

Pew Research Center. (2025, October 15). *How people around the world view AI*. Retrieved from https://www.pewresearch.org/global/2025/10/15/how-people-around-the-world-view-ai/

Price, G. (2024) Japanese University EFL Student Insights on the Emerging ChatGPT Phenomenon, *International Journal of Education, Learning and Development*, 12 (5), 83-100. https://doi.org/10.37745/ijeld.2013/vol12n583100

Price, G., & Sakellarios, M. (2023). The Effectiveness of Free Software for Detecting AI-Generated Writing. *International Journal of Teaching, Learning and Education*. 2 (6). 31-38. https://dx.doi.org/10.22161/ijtle.2.6.4

Rettinger, D. A., & Kramer, Y. (2009). Situational and personal causes of student cheating. Research in Higher Education, 50(3), 293–313. https://doi.org/10.1007/s11162-008-9116-5

Rinnert, C., & Kobayashi, H. (2005). Borrowing words and ideas: Insights from Japanese L1 writers. *Journal of Asian Pacific Communication (John Benjamins Publishing Co.)*, 15(1), https://doi.org/10.1075/japc.15.1.05rin

Sadoshima, S. (2014). アカデミックライティング教育と情報リテラシー: (情報を再定義) し意見を構築できる学生を育てる [Information literacy and academic writing education: Helping students to process information]. *情報の科学と技術* [*Information Science and Technology]*, 64(1), 22–28. https://doi.org/10.18919/jkg.64.1_22

Saito, N. (2024). *The comparative analysis of national policies for research and education regarding usage of generative AI in Japan, the USA and the UK.* [Unpublished manuscript]. https://hdl.handle.net/10419/302516

Sowden, C. (2005). Plagiarism and the culture of multilingual students in higher education abroad. *ELT Journal*, 59(3), 226–233. https://doi.org/10.1093/elt/cci042

Šupak Smolčić, V., & Bilić-Zulle, L. (2013). Patchwork plagiarism–a jigsaw of stolen puzzle pieces. *Biochemia Medica*, *23*(1), 16-18. https://doi.org/10.11613/BM.2013.004

Teeter, J. (2014). Deconstructing attitudes towards plagiarism of Japanese undergraduates in EFL academic writing classes. *English Language Teaching*, 8(1), 95-109. https://doi.org/10.5539/elt.v8n1p95

University of Tokyo. (2023, April 3). About generative AI tools in classes: Policy on use of AI tools in classes. https://utelecon.adm.u-tokyo.ac.jp/en/docs/ai-tools-in-classes/

University of Tsukuba. (2023, May 11). Basic policy on the use of generative AI at the University of Tsukuba. https://www.tsukuba.ac.jp/en/academics/g-document-fall/pdf/2024fall-generative-AI-en.pdf

Vandehey, M., Michael, A., Diekhoff, G., & LaBeff, E. (2007). College cheating: A twenty-year follow-up and the addition of an honor code. *Journal of College Student Development*, 48(4), 468–480. https://doi.org/10.1353/csd.2007.0043

Welding, L. (2023, March 27). Half of college students say using AI on schoolwork is cheating or plagiarism. *BestColleges*. https://www.bestcolleges.com/research/college-students-ai-tools-survey/

Wheeler, G. (2009). Plagiarism in the Japanese universities: Truly a cultural matter? *Journal of Second Language Writing*, 18(1), 17–29. https://doi.org/10.1016/j.jslw.2008.09.004

Yokohama National University. (2023, June). Use of ChatGPT and other generative AI tools. https://www.ynu.ac.jp/hus/nyushi/30294/detail.html

Yu, H., Glanzer, P. L., & Johnson, B. (2017). Why students cheat: A conceptual framework of personal, contextual, and situational factors. In D. Velliaris (Ed.), *Handbook of research on academic misconduct in higher education* (pp. 35–59). IGI Global. https://doi.org/10.4018/978-1-5225-1610-1.ch002

Mapping the Landscape of Executive Function Assessment in Early Childhood

Fatih Kaynar *
Hacettepe University (Türkiye)

Berrin Akman **
Hacettepe University (Türkiye)

Abstract

Understanding and evaluating executive functions (EF) in early childhood remains a pivotal concern for both academic research and practical implementation. Despite the increasing attention devoted to this domain, existing assessment methods vary significantly in scope, intent, and usability. This research examines EF tools applied to children aged 2 to 6, focusing on studies published between 2000 and 2025. Performance-based tasks, while structured and psychometrically sound, often fall short in capturing behavior in real-life settings. In contrast, rating scales and observational tools offer contextually rich data but frequently lack rigorous validation. Neuropsychological batteries provide reliable and valid results, yet their reliance on expert administration and time-intensive procedures limits their application in everyday contexts. Emerging digital assessments enhance flexibility and efficiency; however, they pose concerns regarding digital equity, cultural appropriateness, and ecological validity. The research concludes that no single instrument adequately captures the multifaceted nature of EF in young children. A blended assessment model—attuned to developmental dynamics and sociocultural variability—may offer a more holistic and pragmatic framework for researchers, educators, and policy actors alike.

Keywords**:** executive functions, early childhood, assessment, working memory, cognitive flexibility, inhibitory control, educational applications, developmental considerations, cultural adaptation.

* ***Fatih Kaynar***, Ministry of National Education, ORCID ID: 0000-0001-5142-2316
e-mail: fatih.kaynar@meb.gov.tr

** ***Prof. Dr. Berrin Akman*****,** Hacettepe University, Faculty of Education, Department of Pre-School Education, ORCID ID: https://orcid.org/0000-0001-5668-4382 e-mail: bakman@hecettepe.edu.tr

Correspondence**:** fatih.kaynar@meb.gov.tr

Introduction

Executive functions (EF) refer to a set of cognitive capacities that support goal-directed behavior through planning, monitoring, and adaptation (Barkley, 2012). During early childhood, these abilities develop rapidly, forming the foundation for both academic learning and social-emotional functioning (Davidson et al., 2006; Hughes & Ensor, 2001). The core dimensions—working memory, inhibition, and cognitive flexibility—are often treated as distinct, though in practice they interact closely (Miyake et al., 2000; Garon et al., 2008). Strong early EF skills have been associated not only with school readiness (Blair & Raver, 2015) but also with long-term academic achievement extending into adolescence (Best et al., 2011; Allan et al., 2014).

This review adopts a mapping approach to explore how EF is assessed in children aged 2 to 6, focusing specifically on how current methods align with developmental demands, educational settings, and cultural-linguistic contexts (Munn et al., 2018; Campbell et al., 2023). In doing so, it aims to trace the scope and distribution of tools in the literature and identify areas where empirical coverage remains limited.

Selecting the right tool to assess EF in young children is rarely straightforward. Teachers, clinicians, and researchers alike face challenges stemming from children's short attention spans, language variability, and the cultural specificity of many assessment tasks (Howard et al., 2021; Baggetta & Alexander, 2016). The sheer number of available instruments can further complicate decision-making, contributing to fragmentation across studies (Toplak et al., 2013). Against this backdrop, the current review seeks to classify EF measurement tools into coherent categories and provide a structured overview that may guide future research and practice.

Rationale, Review Objectives, and Structure

The assessment of executive functions (EF) in early childhood has garnered increasing empirical attention over the past two decades, reflecting the growing recognition of EF as a foundational mechanism for cognitive, emotional, and behavioral regulation during the preschool years (Diamond, 2013; Blair, 2016). However, the literature is highly heterogeneous. EF assessment tools diverge not only in format—ranging from direct performance tasks to behavior rating scales and digital applications—but also in theoretical underpinnings, psychometric rigor, and contextual adaptability (Baggetta & Alexander, 2016; Zelazo & Carlson, 2020). While previous reviews have primarily focused on evaluating psychometric validity or educational relevance (Gioia et al., 2003; Becker et al., 2023), few have undertaken a systematic categorization of these tools within a unified conceptual map.

This mapping review seeks to address that lacuna by offering a taxonomically structured overview of EF assessment approaches for children aged 2 to 6. Rather than appraising study quality, the objective is to chart the landscape of assessment tools, grouping them by methodological genre—performance-based tasks, observational tools, neuropsychological instruments, rating scales, and digital platforms—and to delineate critical gaps related to developmental sensitivity, cultural specificity, and technological accessibility (Howard et al., 2021; Campbell et al., 2023).

By organizing the literature in this way, the review serves three primary purposes: to support practitioners in selecting developmentally and culturally appropriate tools; to aid researchers in identifying underrepresented assessment modalities; and to guide policymakers in recognizing areas where the evidence base remains insufficient for programmatic decision-making (McClelland et al., 2014; Blair & Raver, 2015). The methodological procedures that underlie study selection, categorization, and synthesis are detailed in the subsequent section, following PRISMA 2020 guidelines for transparency and replicability (Page et al., 2021).

Methods

This review follows the structure of a mapping study, rather than a conventional systematic review. Accordingly, it does not include a quality appraisal of the selected literature. Instead, emphasis was placed on the transparency and replicability of the search and selection procedures, consistent with the PRISMA 2020 framework (Page et al., 2021).

A comprehensive search was conducted across five databases—PsycINFO, PubMed, Web of Science, Scopus, and ERIC—to identify studies published between 2000 and 2025 that focus on executive function (EF) assessment in early childhood. The search strategy combined the core terms "executive function" and "early childhood" (including related terms such as "preschool" and "nursery") with modifiers like "measurement," "assessment," "developmental sensitivity," "cultural adaptation," and "technology-based assessment."

Studies were included if they were written in English, involved children aged 2 to 6, and focused on EF assessment through empirical research, systematic reviews, adaptation studies, or meta-analyses. Publications were excluded if the full text was unavailable or if the content did not pertain directly to the preschool population.

The selection process unfolded in several stages. An initial pool of 83 records was retrieved. After removing 12 duplicates, 71 unique entries were screened by title and abstract. Of these, 20 were excluded for not meeting inclusion criteria. The full texts of 51 articles were then reviewed, and 11 were removed due to insufficient relevance in terms of scope or methodology. Ultimately, 40 studies were retained for final analysis.

For analytic purposes, studies were grouped according to five widely recognized EF assessment categories: performance-based tests, behavioral rating scales, observational methods, neuropsychological instruments, and technology-assisted tools. A descriptive synthesis was carried out within this classification framework.

Theoretical Foundations of Executive Functions

Executive functions (EF) encompass a set of cognitive skills essential for goal-oriented behavior, including planning, initiating, monitoring, and adjusting one's actions as needed (Miyake et al., 2000; Diamond, 2013). Among these, working memory, inhibitory control, and cognitive flexibility are typically identified as core components. These skills support not only academic development but also interpersonal relationships, emotional regulation, and broader self-regulatory processes (Blair & Razza, 2007; Denham et al., 2012; Allan et al., 2014).

The period between ages 3 and 6 represents a critical window for EF development, driven by structural and functional maturation within the prefrontal cortex (Garon et al., 2008; Gilmore et al., 2021). Developmental milestones appear sequentially: working memory tends to emerge around age three, followed by inhibitory control, and later, cognitive flexibility, typically consolidating by age five (Diamond, 2013; Doebel, 2020; Neuenschwander et al., 2022).

Rather than operating in isolation, these components often function as a unified construct during early childhood. Empirical work involving factor analysis supports this view, particularly in children aged three to five, where EF presents as a single latent factor (Wiebe et al., 2011; Hughes et al., 2010).

The literature also distinguishes between "cold" EF—those rooted in abstract, decontextualized tasks—and "hot" EF, which emerge in emotionally salient or socially complex situations (Zelazo & Carlson, 2012). The latter are especially relevant for motivation, peer interaction, and real-world adaptation (Carlson, 2005; Denham et al., 2012).

Early identification of EF profiles is increasingly recognized as essential for both diagnostic and educational planning purposes. Difficulties in EF have been linked to neurodevelopmental conditions such as ADHD (Barkley, 2012; Willoughby et al., 2021), while higher EF capacity correlates with early success in foundational academic domains like literacy and mathematics (Traverso et al., 2015; Blair & Razza, 2007).

Supportive strategies for fostering EF include offering consistent routines, clear instructions, and structured learning environments. One-on-one guidance and emotionally safe contexts further enhance these skills (Diamond, 2013; Blair, 2016; Howard et al., 2021). Crucially, interventions that account for individual variation and environmental factors tend to be more effective (McClelland, Acock, & Morrison, 2014; Moffitt et al., 2011).

According to Blair and Raver (2015), EF serves as a robust predictor of school readiness, and its development during the preschool years provides long-term benefits. Measuring EF in this period is not only important for early detection of developmental risks but also for evaluating the impact of educational programs (Zelazo & Carlson, 2020). As Howard and Melhuish (2017) observe, a clearer understanding of EF supports the design of more individualized learning pathways. In line with this, Bull and Lee (2014) emphasize that EF assessments can help anticipate academic difficulties well before they fully manifest in formal schooling.

Performance-Based Tests

Among the most widely used tools for assessing executive function (EF) in early childhood are performance-based tasks. These assessments are structured activities designed to isolate and evaluate specific cognitive processes such as working memory, inhibitory control, and cognitive flexibility (Carlson, 2005; Espy & Cwik, 2004). While often presented in engaging or game-like formats, they follow strict protocols to ensure consistency across administrations.

Despite their popularity, these tasks are not without limitations. Their outcomes can be sensitive to contextual variables, particularly in young children. A child's familiarity with the testing environment, their comfort level with the examiner, or momentary shifts in attention can all influence results. Because these instruments are usually administered in controlled settings, they may not fully reflect a child's functioning in more natural, everyday contexts (Zelazo, 2006; Beck et al., 2011). For this reason, performance data should be interpreted alongside other sources of information to provide a more comprehensive view.

A variety of tasks are commonly employed to assess EF. For cognitive flexibility, the Dimensional Change Card Sort (DCCS) is frequently used. Inhibitory control is often evaluated through simple paradigms like the pencil-tap task or “Simon Says” (Carlson & Moses, 2001). Go/no-go tasks, although basic in structure, place considerable demands on response inhibition. The Bear–Dragon task introduces a playful social element, requiring children to follow or suppress commands depending on the puppet issuing them. Working memory is typically assessed via backward digit recall, though interpretations may vary depending on the child’s verbal proficiency (Arslan & Aydın, 2019). The Head-Toes-Knees-Shoulders (HTKS) task stands out for integrating multiple EF domains in a group-friendly classroom format (McClelland et al., 2014; Cameron et al., 2014).

There is strong evidence supporting the reliability of these instruments. Studies have demonstrated that tasks such as HTKS can produce consistent scores across sessions, age groups, and raters (Becker et al., 2023; Bull et al., 2008; Swanson et al., 2009). However, consistency does not necessarily equate to contextual relevance. While HTKS scores are correlated with academic outcomes, interpretations should consider the broader developmental context (Cameron et al., 2014). As with all EF assessment tools, performance-based measures are most informative when used as part of a broader, multi-method evaluation strategy.

Table 1. *Performance-Based Tests*

Test Name	Age Range	EF Component Measured	Reliability (α)	Validity	Ecological Validity	Advantages	Limitations
DCCS	3–7	Cognitive Flexibility	.85–.91	Good construct validity	Low	Quick, measures flexibility	Requires lab setting
PTT	3–6	Inhibitory Control	.82	Strong internal validity	Moderate	Language-free, brief administration	Instructions may be difficult to understand
Go/No-Go	3–6	Inhibitory Control	.78–.87	Measures response inhibition	Low–Moderate	Clear measurement	Sensitive to motivation
Day–Night Stroop	4–6	Inhibitory Control	.76–.88	Measures conflicting response	Low	Measures mental flexibility	May cause conceptual confusion
Simon Says	3–5	Inhibitory Control	-	Visual-auditory matching	High	Game-based, fun	Hard to standardize
Flanker	4–7	Attention / Inhibitory Control	.81	Valid attention filter	Moderate	Measures stimulus control	Requires tablet
BDST	4–7	Working Memory	.79	Assesses auditory memory	Low	Measures working memory	Depends on verbal skills
HTKS	4–8	Multiple (Inhibitory + Flexibility + Memory)	.89	Related to academic achievement	High	Can be used in classroom	Commands may be challenging
HTKS-R	4–8	Multiple (Inhibitory + Flexibility + Memory)	.91	Improved predictive power	High	Increased difficulty levels	May be time-consuming

Behavioral Rating Scales

Rather than assessing executive function (EF) through direct interaction with the child, behavioral rating scales rely on adult observations to infer EF-related abilities. Typically, a caregiver or teacher familiar with the child evaluates behaviors such as sustained attention, impulse control, and emotional regulation across everyday situations (Gioia et al., 2003; Barkley, 2012). One widely recognized tool in this category is the Behavior Rating Inventory of Executive Function–Preschool version (BRIEF-P), which gathers structured feedback from both parents and teachers to generate a profile of EF strengths and weaknesses (Gioia et al., 2003).

These scales offer a valuable counterbalance to performance-based assessments. Because they capture how children behave in real-world contexts—rather than under artificial test conditions—they provide insight into how EF skills manifest in daily routines. Moreover, when multiple informants contribute data, a more nuanced and contextually diverse picture of the child's functioning can emerge (Barnes et al., 2018b; Backer-Grøndahl et al., 2016).

Nonetheless, several limitations accompany this approach. Observational reports are inherently subjective and can vary depending on the rater's expectations, experiences, or familiarity with the child. Two adults may interpret the same behavior differently, especially if the child displays externalizing difficulties or inconsistent behavior across settings (Beck et al., 2011; Barnes et al., 2018b). For this reason, scores derived from behavioral scales are best interpreted in conjunction with objective measures and, where possible, triangulated across multiple observers.

Table 2. *Behavioral Rating Scales*

Test Name	Age Range	EF Component Measured	Reliability (α)	Validity	Ecological Validity	Advantages	Limitations
BRIEF-P	2–5	Multiple	.80–.95	Construct & predictive	High	Parent/teacher observation	Potential for bias
CHEXI	4–12	Multiple	.72–.90	Construct validity	High	Short and understandable	High subjectivity
CBQ	3–7	Inhibitory Control	.70–.89	Construct validity	High	Temperament-based	Indirect measurement
PSRA (Rating)	3–5	Multiple	.85	Strong contextual fit	Moderate–High	Easily applicable in classroom settings	Observer influence

Observational Methods

Observation-based assessments offer an alternative means of evaluating executive function (EF) by focusing on how these skills manifest during everyday experiences. Unlike standardized tests, which isolate cognitive processes in controlled conditions, observational methods aim to capture how EF operates in real time—often within familiar environments like classrooms or play settings (Raver et al., 2011). In these contexts, educators or researchers record behaviors such as sustained attention, impulse control, and problem-solving as they naturally occur during routine activities.

One commonly cited example is the Child Behavior Rating Scale, developed by Bronson et al. (1995), which evaluates children's attention and regulation skills in classroom settings. Another approach, the Executive Function Mapping Protocol created by Bailey and colleagues (2018), uses a structured coding system to identify how various classroom interactions elicit EF-related behaviors.

Because these methods document spontaneous actions in realistic environments, they can reveal nuanced aspects of children's self-regulation that might not emerge during one-on-one testing sessions (Cameron et al., 2021; Ahmed et al., 2021). At the same time, this naturalism introduces challenges. Ratings can vary depending on the observer's interpretations, and differences in classroom dynamics or instructional style can influence how EF skills are expressed (Baggetta & Alexander, 2016). Moreover, observational data can be difficult to standardize since no two settings or time points are exactly alike. For this reason, multiple observations or assessments by different raters are often recommended to increase the reliability and interpretability of the results.

Table 3. *Observational Assessment Scales*

Test Name	Age Range	EF Component Measured	Reliability (α)	Validity	Ecological Validity	Advantages	Limitations
PSRA–O	3–5	Multiple	.83	Validity of natural behavior	High	Applied in real classroom environments	Difficult to standardize
HTKS–O	4–6	Multiple	.86	Predictive validity	High	Game-based observation	Requires precise coding
McCoy Protocol	3–6	Multiple	.81	Construct validity	High	Analysis via video recording	Time-consuming
Group-Based EF	4–6	Multiple	.78	Group-based observation	High	Classroom-based, realistic	Observer variation
EF Mapping	3–5	Multiple	.79	Field-definition validity	High	Provides detailed analysis	Requires expert-level coding

Neuropsychological Tests

When researchers or clinicians require a more fine-grained understanding of a child's cognitive profile, neuropsychological assessments are often the tool of choice. These tests examine a variety of interrelated domains—such as planning ability, working memory, processing speed, and attentional control—not in isolation, but with an awareness of how these systems overlap and influence one another (Anderson et al., 2008). They are typically administered one-on-one by trained professionals and are most commonly found in diagnostic or academic research settings.

That said, these tools are not always well-suited for younger children. Administering them to preschool-aged participants can be logistically difficult. Many tests are lengthy, require sustained attention, and assume verbal or motor skills that children under five may not reliably demonstrate. Compounding the issue, most available batteries have norms beginning at age five, leaving a gap for younger cohorts (Korkman et al., 2007; Brooks et al., 2009).

Even so, when they are used appropriately, the information they yield can be invaluable. Neuropsychological data can highlight not only areas of concern—such as deficits in inhibitory control or memory—but also strengths that might otherwise go unnoticed. For children with suspected learning disorders or attentional difficulties, this level of detail can support both diagnostic clarity and the formulation of tailored support plans (Anderson et al., 2008; Barkley, 2012). In many cases, these assessments form the backbone of individualized education programs, especially when surface-level observation or checklist ratings fall short (Archibald & Kerns, 1999).

Table 4. *Neuropsychological Tests*

Test Name	Age Range	EF Component Measured	Reliability (α)	Validity	Ecological Validity	Advantages	Limitations
NEPSY-II	5–16	Multiple	.70–.92	Construct & clinical validity	Low–Moderate	Comprehensive test battery	Time-consuming, requires specialist
CANTAB	4–16	Multiple	.75–.90	Neuropsychological validity	Low	Tablet-based, detailed	Requires equipment and time
TMT–P	4–6	Cognitive Flexibility	.74	Planning/attention validity	Low	Adapted for children	Slow pace, attention difficulties

Technology-Based Assessment Tools

As digital technologies become more integrated into educational settings, interest in assessing executive function (EF) through digital means has grown. Many of these tools present tasks on tablets or computers, allowing children to respond by tapping or interacting with visual prompts. Some of these assessments are digitized versions of traditional paper-based tasks, while others offer adaptive features that adjust task difficulty in real time based on a child's performance (Akshoomoff et al., 2014; Weintraub et al., 2013).

The NIH Toolbox, for instance, includes a touchscreen-based EF battery suitable for children aged three and above (Zelazo et al., 2013). Other platforms like EF Touch use playful formats to measure skills such as cognitive flexibility and sustained attention in younger age groups (Willoughby et al., 2010; Howard et al., 2021).

There are several reasons why technology-based assessments are considered promising. Responses are captured in real time, scoring is automated, and visual feedback helps keep children

engaged. Tasks can also be shortened or extended depending on the child's pace. In addition, the language and imagery used in these tools can be localized, enabling broader cultural relevance and norm development (Akshoomoff et al., 2014).

Even so, access to appropriate devices and internet infrastructure remains uneven—particularly in under-resourced schools. For some children, especially those unfamiliar with touchscreen devices, performance may reflect comfort with technology more than cognitive ability. In these cases, high scores might not translate to real-world executive function (Zelazo, 2006). For this reason, questions remain about the ecological validity of such assessments. Moreover, children's screen habits and attention variability introduce further complexity in interpreting results.

Still, preliminary findings suggest that, when applied thoughtfully, digital tools can complement traditional methods and help capture dimensions of EF that are harder to observe in structured testing environments (Zelazo et al., 2013).

Table 5. *Technology-Based Assessment Tools*

Test Name	Age Range	EF Component Measured	Reliability (α)	Validity	Ecological Validity	Advantages	Limitations
MEFS	2–7	Multiple	.90	Construct & predictive validity	Moderate	Short tablet-based administration	Requires screen interaction
NIH Toolbox	3+	Multiple	.88–.94	Construct validity	Low–Moderate	Standardized, digital, flexible	Hardware dependency
EF Touch	3–6	Multiple	.85–.92	Application validity	Moderate	Game-based, adaptive	Limited attention span

Comparative Analysis and Evaluation of Methods

No single method can fully capture the complexity of executive function in early childhood. Each one—whether based on performance, observation, or rating—offers a partial view. Some are strong in structure, others in context. But they all come with trade-offs that matter.

Performance-based tasks, for example, are designed to measure specific cognitive processes under controlled conditions. They often show strong internal validity, but children's responses can shift based on how they feel that day or how familiar they are with the testing environment. The lab setting itself can be part of the problem. It's too far removed from everyday life to accurately reflect how EF functions in natural contexts.

Behavioral scales move in a different direction: they rely on how others—teachers, parents—perceive the child in daily life. That brings richer ecological information, but also more subjectivity. One adult might see regulation; another might see disobedience. Their expectations shape what they report, which makes consistency difficult to ensure.

Observational tools sit somewhere in between. When used carefully, they let us see EF in action—right as it's happening. But observers are not neutral. Their training, assumptions, even mood can influence their interpretations. And since no two classroom moments are identical, standardization is difficult.

Neuropsychological assessments are detailed and often precise. They are valuable for diagnostic purposes, especially with older children. But they take time, require expert administration, and are not always feasible with very young children—especially when age-appropriate norms are lacking.

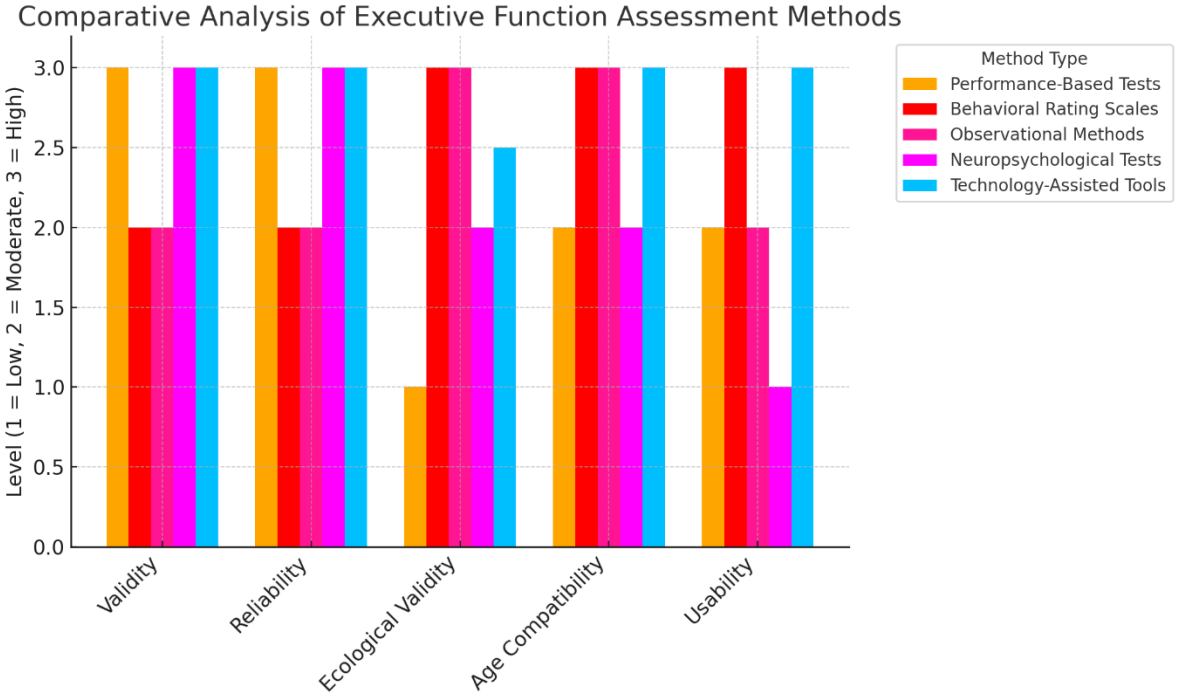
Digital tools introduce a different blend. Touchscreens, adaptive difficulty, real-time scoring—they're efficient and often engaging. Tests like EF Touch and the NIH Toolbox can be applied across a broad age range. However, access to technology and children's familiarity with digital platforms vary significantly across contexts. While they expand potential access, they may also introduce new inequities.

Each method has its strengths and limitations. A combined approach—balancing what each reveals with what each misses—offers the most robust and context-sensitive understanding of EF in young children.

Table 6. *Comparison of Methods.*

Method Type	Validity	Reliability	Ecological Validity	Age Compatibility	Usability
Performance-Based Tests	High	High	Low	Moderate	Moderate
Behavioral Rating Scales	Moderate	Moderate	High	High	High
Observational Methods	Moderate	Moderate	High	High	Moderate
Neuropsychological Tests	High	High	Moderate	Moderate	Low
Technology-Assisted Tools	High	High	Moderate–High	High	High

Figure 1. *Comparison of Methods.*



Developmental Sensitivity and Cultural Adaptation

Developmental Sensitivity of Assessment Approaches

Executive function (EF) abilities develop rapidly in the early years, and children's individual differences often emerge earlier than expected. That's why choosing tools that are developmentally sensitive is not just ideal—it's essential (Carlson & Zelazo, 2014). But this isn't only about age-appropriate difficulty levels. It's also about how well a task fits with a child's attention span, language ability, and even emotional readiness (Beck et al., 2011).

Consider what happens when a test is too easy or too hard: it fails to capture meaningful differences between children. Floor and ceiling effects, especially in young populations, limit our ability to see variation where it really matters (Zelazo, 2013). And if the language used in test instructions is too complex, children may struggle—not because of poor executive function, but simply because they don't understand what's being asked. That's why for younger groups, especially those aged 3 to 4, simple, visual cues can make all the difference (Hughes & Graham, 2002). It also matters whether the task actually reflects the kinds of situations children regularly face. Developmental sensitivity includes how relatable or motivating the activity is. Some assessments, like the BRIEF-P, help in this regard by providing norms tailored to age groups, allowing educators and researchers to compare like with like (Gioia et al., 2000; Thorell & Nyberg, 2008).

Table 7. *Developmental Sensitivity of Methods.*

Assessment Approach	Age Compatibility	Adaptability	Floor/Ceiling Risk	Longitudinal Suitability	Description
Performance-Based Tests	Moderate	Low	Moderate	Limited	Some versions are age-adapted for specific developmental stages
Behavioral Rating Scales	High	Low	Low	Moderate	Norms exist, but longitudinal tracking capacity is limited
Observational Methods	High	Moderate	Low	Moderate	Enables natural behavior observation in real contexts
Neuropsychological Tests	Moderate	Low	High	Low	Complex structure may limit applicability in young children
Technology-Assisted Tools	High	High	Low	High	Tools like MEFS and NIH Toolbox are age-adaptive and scalable

Figure 2. *Characteristics of Different Executive Function Approaches.*

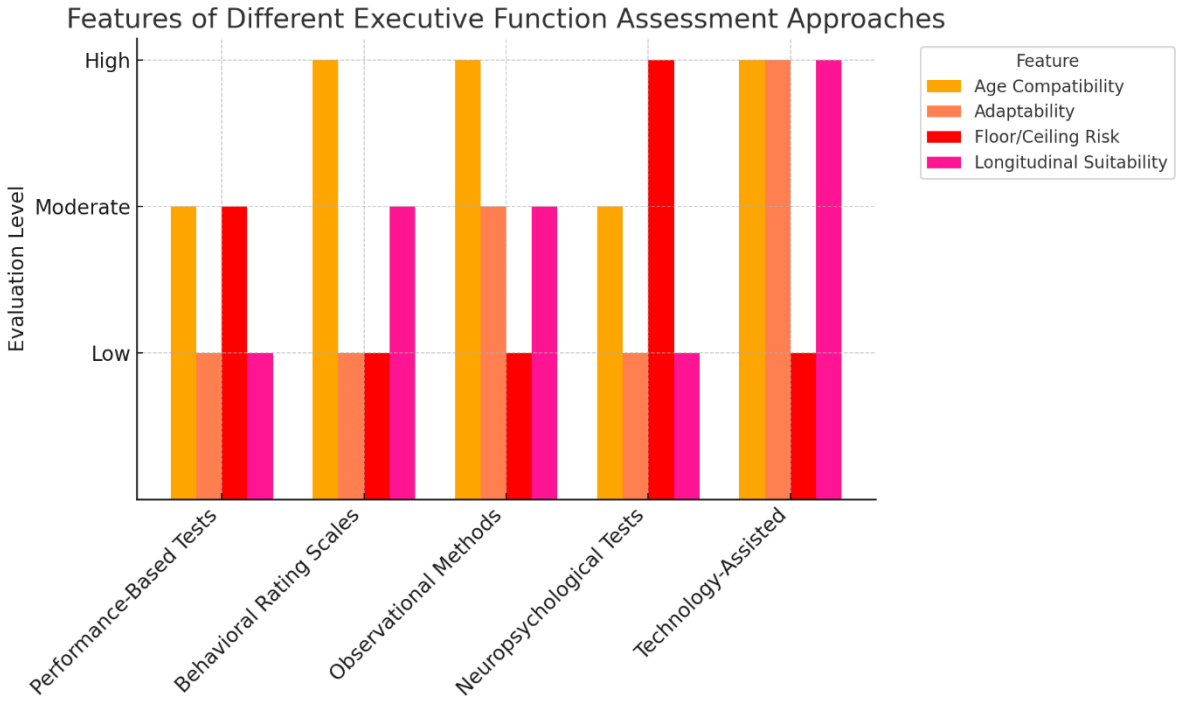


Table 8. *Developmental Sensitivity Features of Specific Tools.*

Assessment Tool	Age Range	Floor/Ceiling Risk	Adaptability	Norm Data	Contextual Fit
DCCS	3–7	Moderate	Available	Available	Moderate
HTKS	4–8	Low	Partial	Available	High
BRIEF-P	2–5	Moderate	Not Available	Available	High
MEFS	2–7	Low	High	Available	High
NIH Toolbox	3–85	Low	High	Extensive	Moderate

Influence of Cultural and Linguistic Factors

How EF is expressed—and how it's interpreted—can vary a great deal across cultures. A behavior viewed as inattentive in one context might be perfectly acceptable in another. Similarly, what looks like a lack of inhibition could simply reflect a child raised in a culture that values assertiveness over obedience (Chevalier et al., 2022; Keller et al., 2013). This is why assessments developed in one region or language often don't travel well without modification. Translating the test isn't enough. The images used, the situations described, and even the behavior expected must make sense within the child's lived reality (Nilsen, 2017; Luciana et al., 2018). Something as basic as a reference to snow or eye contact can mean different things—or nothing at all—depending on the cultural setting (Oh & Lewis, 2008). To adapt assessments responsibly, cultural validation is key. This might involve translation/back-translation, piloting the tool with a sample group, or revising items after expert review (International Test Commission, 2017). Even then, observation-based tools pose their own risks. If a teacher or examiner interprets behavior through their own cultural lens, misjudgments can happen—especially when that behavior doesn't match mainstream expectations (Rogoff et al., 2003; Rogoff et al., 2024).

Educational Assessment of Executive Functions

In recent years, executive functions (EF) have gained prominence in discussions of early learning, not merely as cognitive capacities but as embedded, emergent traits shaped by experience. Especially in early childhood, EF—spanning inhibition, working memory, and cognitive flexibility—unfolds within social routines and guided interactions rather than through isolated tasks (Blair, 2016). While much of the literature treats EF as a fixed skillset to be measured, a more useful view may be that of a developmental orientation: a framework for understanding how children respond to structured environments, manage impulses, and adapt to group expectations (Cartwright, 2012; Diamond & Lee, 2011).

Classroom Observations

Rather than relying solely on tests administered in controlled settings, educators increasingly turn to classroom-based observations to evaluate EF. The logic is intuitive: behaviors such as persistence, shifting between tasks, or managing turn-taking are more readily observed in real time, during authentic learning episodes. For teachers and researchers alike, this offers valuable context that is often missing from abstract testing environments (Gathercole et al., 2006; McCoy et al., 2022).

Structured Observation Tools

Even so, raw observation can be inconsistent without a guiding framework. Tools like the Work Sampling System (WSS) and Child Behavior Rating Scale (CBRS) offer standardized formats for documenting how children initiate, sustain, and regulate engagement across a school day (Bronson

et al., 1995; Meisels et al., 2010). While not exhaustive, these frameworks help capture the nuances of EF development in ways that more static instruments may overlook.

Adaptation of Performance-Based EF Tasks

Interestingly, many tasks originally designed for clinical testing have been reimagined as classroom games. One such example, the HTKS task, integrates bodily movement with verbal inhibition, making it ideal for assessing regulatory control in group settings (McClelland et al., 2014). These adaptations, while simplified, align more closely with the everyday contexts in which EF is actually needed—and used. When children regulate behavior among peers, the skills are on display in ways no lab could simulate.

Executive Functions and Academic Achievement

It is tempting to view EF as a hidden architecture for learning—and in many ways, it is. Research linking EF to early academic performance is both consistent and compelling. Working memory, for example, supports multi-step problem-solving. Inhibitory control helps sustain attention and resist distractions. Flexibility allows children to revise approaches or pivot when plans falter. These are not optional traits in classrooms; they are essential, especially in language-rich or math-intensive environments (Swanson et al., 2009; Blair & Razza, 2007).

Empirical studies bear this out. Children with stronger self-regulatory behavior on tasks like HTKS tend to outperform peers in early literacy and math (Ponitz et al., 2009). The correlations, though not absolute, are robust—often exceeding those between academic outcomes and IQ alone (Alloway & Alloway, 2010). Yet, while numbers suggest strength, they cannot capture the situational variability of a child's functioning from one setting to another.

Figure 3. *Executive Function Components and Academic Achievement.*

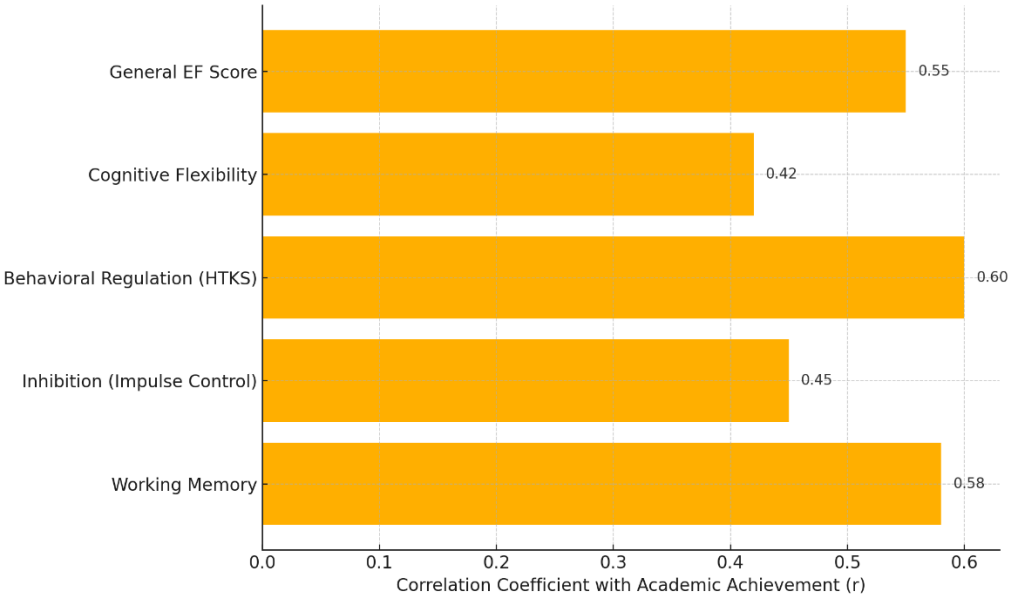


Table 9. *Executive Function–Academic Domain Correlations.*

Executive Function Component	Academic Domain	Correlation Coefficient (r)	Source
Working Memory	Math/Reading	.52–.63	Alloway & Alloway (2010)
Inhibition	Math/Reading	.40–.50	Blair & Razza (2007)
Behavioral Regulation (HTKS)	Reading/Overall Achievement	≈ .60	Ponitz et al. (2009)
Cognitive Flexibility	Reading/Writing	≈ .40+	Cartwright (2012)
All Executive Function Components	Reading & Math	.45–.60	Jacob & Parkinson (2015)

Interventions and Individualized Education Programs

Where EF scores prove especially powerful is in shaping tailored supports. Programs like PATHS and Tools of the Mind not only assess but aim to build EF capacities through developmentally appropriate activities (Diamond et al., 2007; Raver et al., 2011). For children requiring Individualized Education Programs (IEPs), nuanced EF profiles can inform classroom accommodations and instructional pacing. It is critical, however, that these profiles not become static labels. What they offer is not diagnosis but direction—guidance for intervention, not a ceiling on growth (Ylvisaker & Feeney, 2002).

The Role of Parents and Teachers

EF doesn't develop in isolation. Its cultivation depends on consistent interactions with responsive adults—parents, caregivers, educators. Families who understand their child's EF profile can implement routines that reduce unpredictability and support gradual autonomy (Bierman et al., 2008). Meanwhile, teachers benefit from professional learning opportunities that move beyond theory. Training that focuses on identifying EF in action—during transitions, peer conflicts, or collaborative tasks—can shift the entire tone of a classroom (Durlak et al., 2011; McClelland et al., 2014). Ultimately, when adults on all fronts recognize EF not just as a trait but as a teachable skill, their role expands: they become architects of children's regulatory growth.

Discussion, Conclusion, and Recommendations

This review set out to map the range of methods used to assess executive functions (EF) in early childhood and to understand how these approaches align—or fail to align—with developmental, cultural, and educational contexts. Across the reviewed literature, a clear methodological diversity becomes apparent. Some tools, such as performance-based assessments and behavioral rating scales, are widely applied, whereas others—particularly culturally adapted instruments and technology-supported formats—remain underrepresented (Munn et al., 2018; Campbell et al., 2023).

Rather than comparing intervention outcomes or measuring effectiveness, this mapping review aimed to clarify the current landscape of EF assessment. The classification of approaches developed here offers a basis for identifying strengths, limitations, and future research directions. Notably, areas such as ecological validity, cultural relevance, and the broader use of digital technologies require further exploration (Howard et al., 2021; Zelazo et al., 2013).

Figure 4. Comparative Evaluation of Executive Function Assessment Methods

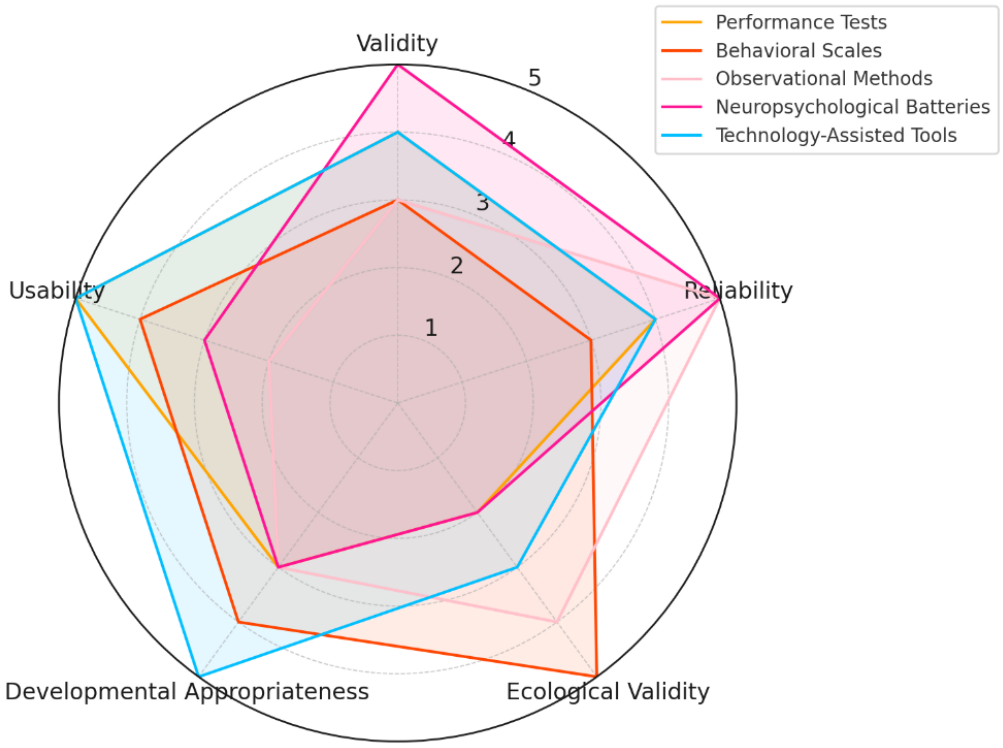


Figure 4 summarizes the comparative strengths and weaknesses across five main categories: performance-based tasks, behavioral rating scales, observational methods, neuropsychological tests, and technology-based tools. Each brings distinct advantages and limitations. For example, neuropsychological tests provide high psychometric validity but are often impractical due to time demands and the need for expert administration. Behavioral rating scales are more accessible and better reflect real-world behaviors, though they rely heavily on subjective reporting, which may lead to bias or inconsistency.

Performance tasks strike a balance between validity and standardization but may not fully reflect children's everyday behaviors. Observational methods offer developmentally rich data but face challenges in standardization and inter-rater reliability. Technology-based tools are scalable and flexible but depend on access to infrastructure and digital fluency—resources not equally distributed across settings.

Overall, the findings point to the value of integrated approaches. No single method captures the full scope of EF in early childhood. By combining insights from multiple sources—structured tasks, rating scales, observations—researchers and practitioners can enhance both the reliability and contextual relevance of assessment outcomes (Diamond, 2013; Baggetta & Alexander, 2016).

Looking ahead, multi-method, multi-informant strategies that are sensitive to culture, development, and context will be key. Assessments that mirror children’s lived experiences—both in home and school environments—are more likely to produce equitable, meaningful, and useful data.

To truly capture the complexity of EF in early childhood, assessments must go beyond technical precision. What is needed is a holistic, flexible approach—one that reflects the diversity of children’s realities and supports a deeper understanding of how they learn, adapt, and grow.

Conflicts of Interest: No conflict of interest has been declared by the authors.

Funding Details: This study was not funded by any organization.

Ethical Statement: This study is a review paper which does not require ethics committee approval.

Credit Author Statement: *Fatih Kaynar*: Conceptualization, Investigation, Writing – original draft & *Berrin Akman*: Methodology, Resources, Validation, Review & Editing.

References

Ahmed, I., Steyer, L., Suntheimer, N. M., & Wolf, S. (2022). Directly assessed and adult-reported executive functions: Associations with academic skills in Ghana. *Journal of Applied Developmental Psychology, 79*, 101379. https://doi.org/10.1016/j.appdev.2022.101379

Ahmed, S. F., Grammer, J., & Morrison, F. J. (2021). Cognition in context: Validating group-based executive function assessments in young children. *Journal of Experimental Child Psychology*, 202, 105005. https://doi.org/10.1016/j.jecp.2020.105005

Akshoomoff, N., Beaumont, J. L., Bauer, P. J., Dikmen, S. S., Gershon, R. C., Mungas, D., ... Heaton, R. K. (2014). NIH Toolbox Cognition Battery (CB): Composite scores of crystallized, fluid, and total cognition. *Monographs of the Society for Research in Child Development, 79*(4), 119–132. https://doi.org/10.1111/mono.12132

Allan, N. P., Hume, L. E., Allan, D. M., Farrington, A. L., & Lonigan, C. J. (2014). Relations between inhibitory control and the development of academic skills in preschool and kindergarten: A meta-analysis. *Developmental Psychology, 50(*10), 2368–2379. https://doi.org/10.1037/a0037493

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*(1), 20–29. https://doi.org/10.1016/j.jecp.2009.11.003

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2008). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 79*(3), 475–490. https://doi.org/10.1111/j.1467-8624.2008.01142

Anderson, P., Anderson, V., Northam, E., Jacobs, R., & Mikiewicz, O. (2008). Development of executive functions through late childhood and adolescence in an Australian sample. *Developmental Neuropsychology, 33*(1), 1–28. https://doi.org/10.1080/87565640802404572

Anderson, V., Anderson, P., Northam, E., Jacobs, R., & Mikiewicz, O. (2008). *Developmental neuropsychology: A clinical approach*. Psychology Press.

Archibald, S. J., & Kerns, K. A. (1999). Identification and description of new tests of executive functioning in children. *Child Neuropsychology, 5*(2), 115–129. https://doi.org/10.1076/chin.5.2.115.3167

Archibald, S. J., & Kerns, K. A. (1999). Identification and description of new tests of executive functioning in children. *Child Neuropsychology, 5* (2), 115–129. https://doi.org/10.1076/chin.5.2.115.3167

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. International Journal of Social Research Methodology, 8(1), 19–32. https://doi.org/10.1080/1364557032000119616

Arslan, A., & Aydın, B. (2019). [Validity and reliability of the numerical auditory feedback test for evaluating executive functions in the preschool period]. Türk Psikolojik Danışma ve Rehberlik Dergisi, 9(53), 22–35.

Arslan, A., & Aydın, B. (2019). Executive function in preschool children: The role of socioeconomic status and parenting styles. Early Child Development and Care, 189(9), 1397–1410. https://doi.org/10.1080/03004430.2017.1371196

Backer-Grøndahl, A., Nærde, A., & Ulleberg, P. (2016). Measuring effortful control using the Children's Behavior Questionnaire–Very Short Form: Modeling matters. Journal of Personality Assessment, 98(6), 581–590. https://doi.org/10.1080/00223891.2015.1056303

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

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

Bailey, R., Jones, S. M., & Partee, A. (2018). Executive Function Mapping Project Measures Compendium: A resource for selecting measures related to executive function and other regulation-related skills in early childhood (OPRE Report No. 2018-59). Office of Planning, Research, and Evaluation.

Barkley, R. A. (2012). Executive functions: What they are, how they work, and why they evolved. Guilford Press.

Barnes, J. J., Howard, J., Green, D., & Astle, D. E. (2018). Parental perspectives on neurocognitive executive function skills. Frontiers in Psychology, 9, 2092. https://doi.org/10.3389/fpsyg.2018.02092

Barnes, J. J., Nobre, A. C., Woolrich, M. W., Baker, K., & Astle, D. E. (2018). The neural and genetic basis of working memory development in childhood. Developmental Cognitive Neuroscience, 30, 1–12. https://doi.org/10.1016/j.dcn.2017.07.003

Beck, D. M., Espy, K. A., & Wiebe, S. A. (2011). A theoretical and empirical review of executive function. Frontiers in Psychology, 2, 163. https://doi.org/10.3389/fpsyg.2011.00163

Becker, D. R., Fernandes, V. R., & McClelland, M. M. (2023). Head-Toes-Knees-Shoulders task and EF in two samples of adolescents in Brazil and United States. Frontiers in Psychology, 14, 1149053. https://doi.org/10.3389/fpsyg.2023.1149053

Becker, D. R., McClelland, M. M., Tominey, S. L., & Koval, C. (2023). Longitudinal links between behavioral regulation and academic achievement. Early Childhood Research Quarterly, 64, 1–14. https://doi.org/10.1016/j.ecresq.2023.01.003

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. Learning and Individual Differences, 21(4), 327–336. https://doi.org/10.1016/j.lindif.2011.01.007

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., ... Gill, S. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI

program. Child Development, 79(6), 1802–1817. https://doi.org/10.1111/j.1467-8624.2008.01227.x

Blair, C. (2016). Developmental science and executive function. Current Directions in Psychological Science, 25(1), 3–7. https://doi.org/10.1177/0963721415622634

Blair, C., & Raver, C. C. (2015). School readiness and self-regulation: A developmental psychobiological approach. Annual Review of Psychology, 66, 711–731. https://doi.org/10.1146/annurev-psych-010814-015221

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. Child Development, 78(2), 647–663. https://doi.org/10.1111/j.1467-8624.2007.01019.x

Blair, C., & Raver, C. C. (2015). School readiness and self-regulation: A developmental psychobiological approach. Annual Review of Psychology, 66, 711–731. https://doi.org/10.1146/annurev-psych-010814-015221

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. Child Development, 78(2), 647–663. https://doi.org/10.1111/j.1467-8624.2007.01019.x

Bronson, M. B., Goodson, B. D., Layzer, J. I., & Love, J. M. (1995). Child Behavior Rating Scale. Abt Associates.

Brooks, B. L., Sherman, E. M., & Strauss, E. (2009). NEPSY-II: A developmental neuropsychological assessment. Child Neuropsychology, 15(6), 564–578. https://doi.org/10.1080/09297040903146966

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. Developmental Neuropsychology, 33(3), 205–228. https://doi.org/10.1080/87565640801982312

Cameron, C. E., McClelland, M. M., Duncan, R., Bowles, R. P., Acock, A. C., & Miao, A. (2014). Predictors of early growth in academic achievement: The Head-Toes-Knees-Shoulders task. Frontiers in Psychology, 5, 599. https://doi.org/10.3389/fpsyg.2014.00599

Cameron, C. E., McClelland, M. M., & Gonzales, C. R. (2021). The Head-Toes-Knees-Shoulders revised: Links to academic outcomes and measures of EF in young children. Frontiers in Psychology, 12, 721846. https://doi.org/10.3389/fpsyg.2021.721846

Campbell, C. E., Wang, Y., & Howard, S. J. (2023). Developmental trajectories in executive function measurement. Child Development Perspectives, 17(1), 24–31. https://doi.org/10.1111/cdep.12461

Campbell, F., Tricco, A. C., Hersi, M., & Moher, D. (2023). Mapping reviews, scoping reviews, and evidence and gap maps (EGMs): The same but different—the “Big Picture” review family. Systematic Reviews, 12(1), 45. https://doi.org/10.1186/s13643-023-02178-5

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. Developmental Neuropsychology, 28(2), 595–616. https://doi.org/10.1207/s15326942dn2802_3

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. Child Development, 72(4), 1032–1053. https://doi.org/10.1111/1467-8624.00333

Cartwright, K. B. (2012). Insights from cognitive neuroscience: The importance of executive function for early reading development and education. Early Education and Development, 23(1), 24–36. https://doi.org/10.1080/10409289.2011.615025

Chevalier, N., & Blaye, A. (2009). Dimensional change card sort performance in preschoolers: What's in a label? Journal of Experimental Child Psychology, 103(3), 249–266. https://doi.org/10.1016/j.jecp.2009.01.006

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. Neuropsychologia, 44(11), 2037–2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006

Diamond, A. (2013). Executive functions. Annual Review of Psychology, 64, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. Science, 333(6045), 959–964. https://doi.org/10.1126/science.1204529

Doebel, S. (2020). Rethinking executive function and its development. Perspectives on Psychological Science, 15(4), 942–956. https://doi.org/10.1177/1745691620904771

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. Child Development, 82(1), 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x

Engel de Abreu, P. M. J., Conway, A. R. A., & Gathercole, S. E. (2010). Working memory and fluid intelligence in young children. Intelligence, 38(6), 552–561. https://doi.org/10.1016/j.intell.2010.07.003

Espy, K. A., & Cwik, M. F. (2004). The development of a trail making test in young children: The TRAILS-P. The Clinical Neuropsychologist, 18(3), 411–422. https://doi.org/10.1080/13854040490524277

Espy, K. A., & Cwik, M. F. (2004). The development of executive function: A review of the current conceptualizations and empirical evidence. Developmental Neuropsychology, 27(1), 79–114. https://doi.org/10.1207/s15326942dn2701_4

Fernandez, A. L., & Marcopulos, B. A. (2008). A comparison of normative data for the Trail Making Test from several countries: Equivalence of norms and considerations for interpretation. Scandinavian Journal of Psychology, 49(3), 239–246. https://doi.org/10.1111/j.1467-9450.2008.00637.x

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. Psychological Bulletin, 134(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2006). Working memory skills and educational attainment: Evidence from National Curriculum assessments at 7 and 14 years of age. Applied Cognitive Psychology, 20(1), 1–16. https://doi.org/10.1002/acp.934

Gioia, D. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2003). Behavior Rating Inventory of Executive Function–Preschool Version (BRIEF-P). Psychological Assessment Resources.

Howard, S. J., Melhuish, E., & Beauchamp, M. H. (2021). The early development of executive function: A global perspective. Developmental Science, 24(1), e12968. https://doi.org/10.1111/desc.12968

Howard, S. J., Melhuish, E., & Martin, A. J. (2021). An integrated model of executive functioning and self-regulation for school readiness: A theoretical review. Educational Psychology Review, 33, 1–34. https://doi.org/10.1007/s10648-020-09540-4

Howard, S. J., Okely, A. D., & Ellis, Y. G. (2021). Executive function assessment in early childhood: A review of methodological and conceptual considerations. Developmental Review, 62, 100983. https://doi.org/10.1016/j.dr.2021.100983

Hughes, C., & Ensor, R. (2001). Individual differences in theory of mind, executive function, and social understanding in the preschool years. Child Development, 72(3), 878–896. https://doi.org/10.1111/1467-8624.00320

Hughes, C., & Graham, A. (2002). Measuring executive functions in childhood: Problems and solutions? Child and Adolescent Mental Health, 7(3), 131–142. https://doi.org/10.1111/1475-3588.00023

Korkman, M., Kirk, U., & Kemp, S. (2007). NEPSY-II: Clinical and interpretive manual. Psychological Corporation.

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., & Miao, A. (2014). Predictors of early growth in academic achievement: The Head-Toes-Knees-Shoulders task. Frontiers in Psychology, 5, 599. https://doi.org/10.3389/fpsyg.2014.00599

McClelland, M. M., Tominey, S. L., Schmitt, S. A., & Duncan, R. (2014). SEL Interventions in Early Childhood. Future of Children, 27(1), 33–47. https://doi.org/10.1353/foc.2017.0002

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

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Medical Research Methodology, 18, 143. https://doi.org/10.1186/s12874-018-0611-x

Nilsen, E. S. (2017). Cross-cultural considerations in assessing executive function. Journal of Cognitive Development, 18(1), 1–15.

Oh, S., & Lewis, C. (2008). Korean preschoolers' advanced inhibitory control and its relation to other executive skills and mental state understanding. Child Development, 79(1), 80–99. https://doi.org/10.1111/j.1467-8624.2007.01112.x

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Metzger, M., & Solomon, B. (2011). CSRP's impact on low-income preschoolers' preacademic skills: Self-regulation as a mediating

mechanism. Child Development, 82(1), 362–378. https://doi.org/10.1111/j.1467-8624.2010.01561.x

Swanson, H. L., Jerman, O., & Zheng, X. (2009). Growth in working memory and mathematical problem solving. Journal of Educational Psychology, 101(2), 296–311. https://doi.org/10.1037/a0013841

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

Weintraub, S., Bauer, P. J., Zelazo, P. D., Wallner-Allen, K., Dikmen, S. S., Heaton, R. K., ... & Gershon, R. C. (2013). Cognition assessment using the NIH Toolbox. Neurology, 80(11 Suppl 3), S54–S64. https://doi.org/10.1212/WNL.0b013e3182872ded

Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M. T. (2010). The measurement of executive function at age 3 years: Psychometric properties and criterion validity of a new battery of tasks. Psychological Assessment, 22(2), 306–317. https://doi.org/10.1037/a0018708

Ylvisaker, M., & Feeney, T. (2002). Executive functions, self-regulation, and learned optimism in children with traumatic brain injury and children with typical development. School Psychology Quarterly, 17(4), 378–408. https://doi.org/10.1521/scpq.17.4.378.20885

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. Nature Protocols, 1(1), 297–301. https://doi.org/10.1038/nprot.2006.46

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. Monographs of the Society for Research in Child Development, 78(4), 16–33. https://doi.org/10.1111/mono.12032

Zelazo, P. D., & Carlson, S. M. (2020). The neurodevelopment of executive function skills: Implications for academic achievement gaps. Psychology and Neuroscience, 13(1), 1–19. https://doi.org/10.1037/pne0000176

Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. Child Development Perspectives, 6(4), 354–360. https://doi.org/10.1111/j.1750-8606.2012.00246.x

Zelazo, P. D., & Carlson, S. M. (2020). The neurodevelopment of executive function skills: Implications for academic achievement gaps. Psychological Science in the Public Interest, 21 (1), 118–146. https://doi.org/10.1177/1529100620942705