Evaluation of the Psychometric Properties of the Gap-Overlap Task in 10-Month-Old Infants

Cousijn, J.; Hessels, R.S.; Van der Stigchel, S.; Kemner, C.

DOI
10.1111/infa.12185

Publication date
2017

Document Version
Final published version

Published in
Infancy

License
Article 25fa Dutch Copyright Act

INFANCY

THE OFFICIAL JOURNAL OF THE
INTERNATIONAL CONGRESS
OF INFANT STUDIES

Infancy, **22**(4), 571–579, 2017

ISSN: 1525-0008 print / 1532-7078 online
DOI: 10.1111/infa.12185

WILEY Blackwell

Evaluation of the Psychometric Properties of the Gap-Overlap Task in 10-Month-Old Infants

Janna Cousijn
Experimental Psychology
Utrecht University and
Department of Developmental Psychology
University of Amsterdam

Roy S. Hessels
Experimental Psychology
Utrecht University and
Developmental Psychology
Utrecht University

Stefan Van der Stigchel
Experimental Psychology
Utrecht University

Chantal Kemner
Experimental Psychology
Utrecht University and
Developmental Psychology
Utrecht University and
Brain Center Rudolf Magnus
University Medical Centre

Understanding the development of attention is key to understanding cognitive maturation. The gap-overlap task can be administered at all ages and is widely used to study the development of overt visual attention. However, studies using the gap-overlap task report different measures and little is known about the tasks' psychometric properties, especially in infants. We tested the 1-week test–retest reliability of two frequently used gap-overlap measures of attentional disengagement in 10-month-old infants; the gap effect as measured by the difference between the gap and overlap condition and the gap

Correspondence should be sent to Janna Cousijn, Department of Developmental Psychology, University of Amsterdam, P.O. box 15916, 1001NK Amsterdam, The Netherlands. E-mail: j.cousijn@gmail.com

effect as measured by the difference between the gap and baseline condition. Sixty-seven infants performed the gap-overlap task twice, of which 45 infants had sufficient data quality for further analyses. Test–retest reliability of the overlap-gap gap effect was higher (r = .50) than the baseline-gap gap effect (r = .29). Moreover, the shared variance between overlap and baseline saccadic reaction times was moderate to high across sessions. In light of these results and the methodological challenges and limitations of infant research, we consider the overlap-gap gap effect to be a good measure to study the development of attentional disengagement in infants and suggest the exclusion of the baseline condition in future studies.

The gap-overlap task is widely used to study the development of overt visual attention across the lifespan (e.g., Crawford et al., 2013; Elsabbagh et al., 2013; Fischer, Gezeck, & Hartnegg, 2000; Mosconi et al., 2009). However, little is known about the psychometric properties of the gap-overlap task, especially in infants. This study therefore aimed to assess test–retest reliability of the gap-overlap task in 10-month-old infants.

A typical trial of the gap-overlap task starts with a central stimulus and after the participant fixates the central stimulus, a peripheral stimulus appears to the left or right of the central stimulus. The participant subsequently makes a saccade toward the peripheral stimulus and the latency to initiate the eye movement from the central to the peripheral stimulus is recorded. There are usually two conditions: gap and overlap. During gap trials, the central stimulus typically disappears 200 ms before the onset of the peripheral stimulus. During overlap trials, the central stimulus remains on screen while the peripheral stimulus is presented. Saccadic reaction times (SRTs) are generally longer during the overlap versus the gap condition because active disengagement from the central stimulus is required during the overlap trials. The so-called gap effect represents the decrease in SRTs during gap versus overlap trials. This gap effect is well replicated and decreases over the course of infancy as the capacity to disengage attention from the central stimulus improves (Elsabbagh et al., 2013; Matsuzawa & Shimojo, 1997; McConnell & Bryson, 2005).

Importantly, different variants of the gap-overlap task have been used in developmental studies and different measures are reported upon. In contrast to most studies in children, adolescents and adults, infant studies often include a baseline condition during which the onset of the peripheral stimulus immediately follows the offset of the central stimulus (e.g., Hood & Atkinson, 1993; Nakagawa & Sukigara, 2013; Wass, Porayska-Pomsta, & Johnson, 2011). Given the fact that only limited testing is possible in infants, it is important to know whether the baseline condition provides additional information, as has been argued earlier (Klein & Foerster, 2001; Reuter-Lorenz, Hughes, & Fendrich, 1991). Another issue is the test–retest reliability of different gap-overlap measures; so far only one study was dedicated to this important question in children and adults (Klein and Fischer (2005). Here we tested the 1-week test–retest reliability of the frequently used gap-overlap measures in 10-month-old infants. We specifically compared the reliability of the gap effect as measured by the difference between the gap and overlap condition and the gap effect as measured by the difference between the gap and baseline condition.

METHODS

Participants

This study was embedded in a larger project on infant cognition (Dynamics of Youth —YOUth cohort; Hessels, Andersson, Hooge, Nyström, & Kemner, 2015). Seventy-seven 10-month-old healthy infants born after 37 weeks of pregnancy with normal hearing and/or vision were recruited via the local municipality, of which 68 performed the gap-overlap task twice (three no-shows, six were crying and/or too tired to start). Insufficient data quality led to a further exclusion of 23 infants (see results on data quality below). The final sample consisted of 45 infants (Table 1). The primary caregiver signed informed consent and received 10 Euro for each test-session. The study was approved by the Medical Ethics Committee of University Medical Center Utrecht (Protocol ID 14-221).

Gap-overlap task

The gap-overlap task was based on Elsabbagh et al. (2013). Trials started with a central clock (2.1° × 2.1°) expanding and contracting (max. size 3.3° × 3.3°) to attract the attention. The central stimulus started spinning 500°/sec after the child fixated it to maintain the infant's attention. A peripheral stimulus (sun, cloud, ball, star, dog, 2.5° × 2.5°, positioned 19° left or right from the central stimulus) appeared 600–700 ms after the infant fixated to the central stimulus. This 100 ms jitter was implemented to decrease anticipatory saccades. The task contained a gap, overlap, and baseline condition. In the gap condition, central stimulus offset was 222 ± 35 ms before peripheral stimulus onset. In the overlap condition, the central and peripheral stimulus remained simultaneously and inanimately on screen. In the baseline condition, peripheral stimulus onset directly followed central stimulus offset. The peripheral

TABLE 1
Sample Characteristics and Gap-Overlap Task Performance

	Session 1	*Session 2*
Demographics		
Final sample (*n*)	45	45
Age in days	304.7 (11.3)	312.0 (11.4)
Gender, % female	58	58
Gap-overlap task performance		
Overall SRT (ms)	295 (38)	280 (42)
Overall included trials (*n*)	38.0 (9.8)	36.8 (11.0)
Gap SRT (ms)	215 (24)	204 (25)
Gap included trials (*n*)	12.3 (3.9)	11.4 (4.0)
Overlap SRT (ms)	387 (79)	361 (83)
Overlap included trials (*n*)	12.6 (3.4)	12.5 (3.9)
Baseline SRT (ms)	283 (39)	277 (45)
Baseline included trials (*n*)	13.0 (3.5)	12.8 (4.1)
Gap-effect: overlap-gap (ms)	172 (76)	157 (82)
Gap effect: baseline-gap (ms)	86 (32)	73 (37)

SRT, Saccadic Reaction Time. Means and standard deviations (SD) are reported if not otherwise indicated. Gap-overlap task contained 60 trials, 20 per condition.

stimulus stayed on screen until the child fixated it or until 1,500 ms elapsed. After a first fixation, the peripheral stimulus either contracted and expanded or spiraled of view for 1,000 ms. This feedback was combined with various sounds (e.g., a car horn, a bell). The gap-overlap task consisted of 20 trials per condition, randomly presented.

Apparatus

The Tobii TX300 eye-tracker (Tobii Technology, Stockholm, Sweden) with an integrated 23-inch monitor (1920 by 1080 pixels; 60 Hz refresh rate) was used to record infants' eye movements. Median measurement precision for all detected fixations was 0.42° root mean square (RMS) noise (SD = 0.20°) in Session 1 and 0.42° RMS noise (SD = 0.21°) in Session 2. The Tobii TX300 ran at 300 Hz and communicated with MATLAB (version R2013a, MathWorks Inc., Natick, MA, USA) and the Psych Toolbox (version 3.0.11; Brainard, 1997) running on a MacBook Pro (OS X 10.9) via the Tobii SDK.

Procedure

Familiarization and positioning (65 cm in front of the eye-tracker) of the infant was performed as is described in Hessels et al. (2015). The system-controlled calibration procedure consisted of a contracting circle that was consecutively presented at all four corners and the center of the screen. Upon full contraction, the point was calibrated, after which it moved to the next point and expanded again. The circles were coupled with sound. If calibration was judged insufficient, the calibration was repeated. The gap-overlap task started after calibration. The infants' looking behavior was followed real-time via a webcam. If the infant lost attention, the experimenter attempted to redirect attention toward the task by playing sounds. The task including calibration lasted 10–15 min.

Data preparation and analyses

Raw position signals from both eyes were first averaged. If gaze position was only available from one eye, that signal was used. Cubic spline interpolation was performed to estimate gaze position for periods of data loss with a duration of <100 ms (Frank, Vul, & Johnson, 2009). The identification by 2-means clustering algorithm was used to extract SRTs (I2MC; Hessels, Niehorster, Kemner, & Hooge, 2016). SRT was defined as the time between target stimulus onset and the first fixation on the target.

Inspection of normality plots, skewness and kurtosis showed that the data distributions were approximately normal. Pearson correlations and intraclass correlations (ICC(A,1); Weir, 2005) were computed to assess test–retest reliability (linear relation and absolute agreement, respectively) of the overall SRTs, and the overlap-gap and baseline-gap gap effect.

RESULTS

Data quality

Infants with <10 included trials in any session were excluded. Trials were excluded when no fixations were detected, the initial fixation was more than 3.6° from the central fixation point, the fixation after target onset was more than 5.5° from the target, or the latency of this fixation was <100 ms or more than 1,000 ms after target onset. This led to the exclusion of 23 infants. In the remaining 45 infants, these criteria led to 36.7% excluded trials in the first session and 38.7% excluded trials in the second session (see Table 1 for included trials per condition). The number of included trials in each session did not correlate with average SRTs in session 1 (r = .1373, p = .368), nor in session 2 (r = −.11653, p = .446).

Fixation latencies over sessions

Infants fixation SRTs were lower during session 2 ($t(44) = 2.82$, $p < .01$), pointing to a typical training effect. SRTs were significantly correlated between sessions (overall: $r = .65$, $p < .001$; Gap: $r = .60$, $p < .001$; Overlap: $r = .54$, $p < .001$; Baseline: $r = .53$, $p < .001$).

Gap effect over sessions

Overlap-gap

In both sessions, a significant difference between the gap and the overlap condition was observed (Session 1: $t(44) = 15.25$, $p < .001$; Session 2: $t(44) = 12.89$, $p < .001$; Figure 1). This difference was similar between sessions ($t(44) = 1.29$, $p = .20$) and significantly correlated ($r = .50$, $p < .001$, Figure 2). ICC(A,1) analysis testing for absolute agreement of this gap effect between sessions indicated significant agreement (ICC = .50, $F(1, 44) = 3.00$, $p < .001$).

Baseline-gap

In both sessions, a significant difference between the gap and the baseline condition was observed (Session 1: $t(44) = 14.37$, $p < .001$; Session 2: $t(44) = 13.15$; $p < .001$; Figure 1). This difference was similar between sessions ($t(44) = .75$; $p = .46$), but was only marginally correlated ($r = .29$; $p = .053$, Figure 2). ICC(A,1) analysis testing for absolute agreement of this gap effect between sessions indicated significant agreement (ICC = .29, $F(1, 44) = 1.80$, $p = .027$).

Comparing the gap effect measures

In both sessions, the difference between the overlap-gap condition was larger than the difference between the baseline-gap condition (Session 1: $t(44) = 9.87$, $p < .001$; Session 2: $t(44) = 8.45$, $p < .001$). Partial correlation analyses showed that baseline and overlap, corrected for gap SRTs correlated significantly (Session 1: $r = .36$, $p = .015$; Session 2: $r = .60$; $p < .001$). Similarly, baseline and gap, corrected for overlap SRTs correlated significantly (Session 1: $r = .59$, $p < .001$; Session 2: $r = .51$, $p < .001$). However, gap and overlap, corrected for baseline SRTs did not correlate (Session 1: $r = .02$; $p = .87$; Session 2 $r = .23$; $p = .14$). This suggests that the baseline condition measures overlapping processes with the gap and overlap condition, but the gap and overlap condition do not.

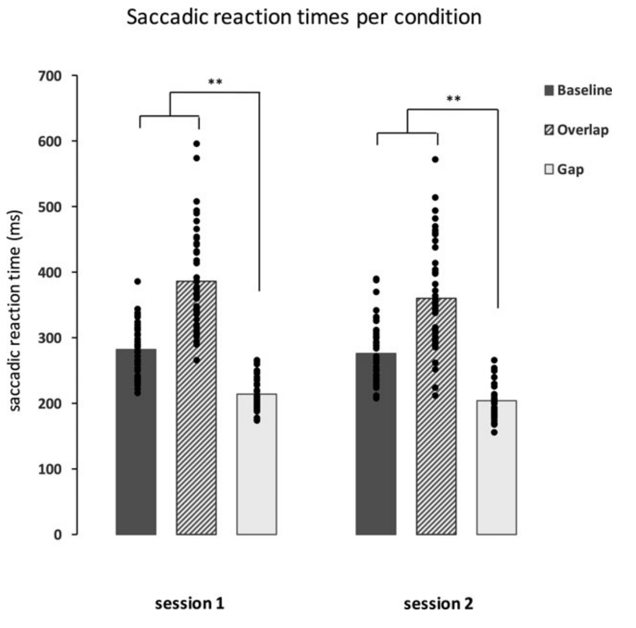

Figure 1 Saccadic reaction times for baseline, overlap and gap trials per session. Bars depict average saccadic reaction time per condition. Black dots indicate individual data points. Gap trails are faster compared to baseline and overlap trials in both sessions. $^{**}p < 0.001$.

DISCUSSION

Studies using the gap-overlap task report different measures and little is known about the tasks' psychometric properties, especially in infants. We therefore assessed test–retest reliability of two commonly used gap effect measures in 10-month-old infants.

The test–retest reliability of the overlap-gap gap effect was $r = .50$ and of the baseline-gap gap effect $r = .29$. Although test–retest reliability scores are generally only acceptable when $r > .70$ (Field, 2013), reliability scores in infant research are lower and only the "best" paradigms may just reach an $r = .70$ (for example see Cristia, Seidl, Singh, & Houston, 2016). In infant eye-tracking, fixation duration appears to have a good test–retest reliability of $r > .70$ (Hessels, Hooge, et al., 2016; Wass & Smith, 2014) but SRTs are generally less reliable. For example, a similar study in 10-month-old infants reported an SRT test–retest reliability of $r = .47$ during a visual search task (Hessels, Hooge, et al., 2016). In light of infant research, we consider the overlap-gap gap effect, but not the baseline-gap gap effect, to be a good measure to study the development of overt visual attention.

Data quality but also measurement time is limited in infants, strongly constraining the number of experimental conditions that can be tested with a single task. In the current study, 36% of the infants needed to be excluded from further analyses and an additional 38% of the trials needed to be excluded in the included infants. Based on the higher test–retest reliability of the overlap-gap gap effect compared to the baseline-gap gap effect, and the limited measurement time in infants we would advise to remove the baseline condition for the infant gap-overlap task [as is already common in versions of the task for older age groups (e.g., Fischer et al., 2000; Goepel, Biehl, Kissler,

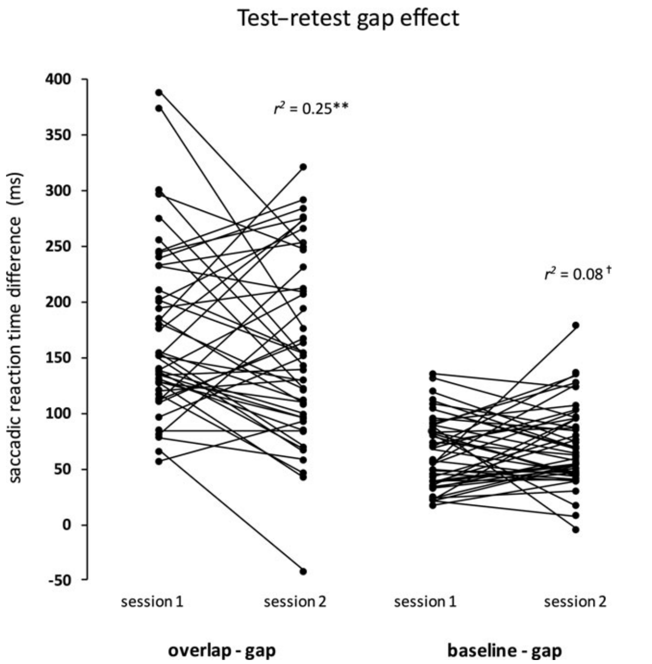

Figure 2 Test–retest reliability of common gap effect measures. Left side of the chart depict the test–retest reliability of the gap effect in milliseconds calculated by subtracting average saccadic reaction times (SRTs) on gap trails from overlap trials. Right side of the chart depict the test–retest reliability of the gap effect in milliseconds calculated by subtracting average SRTs on gap trials from baseline trials. Each line connects the gap effect on session 1 with session 2 for a single infant. $^{**}p < 0.001$, $^{\dagger}p = 0.053$.

& Paul-Jordanov, 2011; Kikuchi et al., 2011; Klein & Foerster, 2001; Mosconi et al., 2009)]. A second argument for removing the baseline condition is the moderate to strong correlation between the baseline and overlap SRTs, after correcting for gap SRTs. This suggests that both measure tap into overlapping aspects of attentional control and are therefore partially redundant.

The better test–retest correlation of the overlap-gap gap effect compared to the baseline-gap gap effect might be explained by the additional processes that are measured by the baseline condition compared to the overlap condition. In the baseline condition, the fixation stimulus is removed at the moment the target is presented. Besides fixation release, this offset also provides a warning signal triggering increased alertness and facilitating disengagement (Klein & Foerster, 2001; Reuter-Lorenz et al., 1991). For this reason, it has previously been argued that the difference between the gap and overlap condition is a cleaner indicator of attentional disengagement compared to the baseline versus gap (Klein & Foerster, 2001; Reuter-Lorenz et al., 1991). Although this cannot directly be verified in the current dataset, this theoretical notion provides a third argument for removing the baseline condition from the infant gap-overlap task in future studies.

While this is the first assessment of test–retest reliability of the gap-overlap task in infants, future studies should investigate if these results hold for other (pre)clinical populations and age groups. This is especially important as reliability may improve with age and disengagement latencies during the gap-overlap task at 14 months, but not at 7 months, were found to be associated with later diagnostic outcome in infants with a familiar risk for Autism (Elsabbagh et al., 2013).

To conclude, we consider the overlap-gap gap effect to be a good measure to study the development of attentional disengagement in infants and suggest the exclusion of the baseline condition in future studies.

ACKNOWLEDGMENTS

We would like to thank Prof. Mark Johnson, Dr. Emily Jones and Dr. Luke Mason, Centre for Brain & Cognitive Development, Birkbeck, University of London for sharing the gap-overlap infant paradigm and for their help in designing and programming the task. Furthermore, we would like to thank all employees of the Kinder Kennis Centrum, Utrecht University for their help during the data collection. This project was supported by the Consortium Individual Development (CID) and by the Utrecht University, through the strategic theme "Dynamics of Youth". CID is funded through the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.001.003 awarded to Chantal Kemner, Utrecht University, The Netherlands).

REFERENCES

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Crawford, T. J., Higham, S., Mayes, J., Dale, M., Shaunak, S., & Lekwuwa, G. (2013). The role of working memory and attentional disengagement on inhibitory control: Effects of aging and Alzheimer's disease. *Age*, *35*, 1637–1650.

Cristia, A., Seidl, A., Singh, L., & Houston, D. (2016). Test-retest reliability in infant speech perception tasks. *Infancy*, *21*, 648–667.

Elsabbagh, M., Fernandes, J., Jane Webb, S., Dawson, G., Charman, T., Johnson, M. H. & British Autism Study of Infant Siblings Team (2013). Disengagement of visual attention in infancy is associated with emerging autism in toddlerhood. *Biological Psychiatry*, *74*, 189–194.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*, 4th ed. London, UK: Sage.

Fischer, B., Gezeck, S., & Hartnegg, K. (2000). On the production and correction of involuntary prosaccades in a gap antisaccade task. *Vision Research*, *40*, 2211–2217.

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*, 160–170.

Goepel, J., Biehl, S. C., Kissler, J., & Paul-Jordanov, I. (2011). Pro- and antisaccades in children elicited by visual and acoustic targets – does modality matter? *BMC Pediatrics*, *11*, 116.

Hessels, R. S., Andersson, R., Hooge, I. T. C., Nyström, M., & Kemner, C. (2015). Consequences of eye color, positioning, and head movement for eye-tracking data quality in infant research. *Infancy*, *20*, 601–633.

Hessels, R. S., Hooge, I. T. C., & Kemner, C. (2016). An in-depth look at saccadic search in infancy. *Journal of Vision*, *16*(8), 10, 11–14.

Hessels, R. S., Niehorster, D. C., Kemner, C., & Hooge, I. T. C. (2016). Noise-robust fixation detection in eye-movement data – Identification by 2-means clustering (I2MC). *Behavior Research Methods*. doi: 10.3758/s13428-016-0822-1

Hood, B. M., & Atkinson, J. (1993). Disengaging visual attention in the infant and adult. *Infant Behavior and Development*, *16*, 405–422.

Kikuchi, Y., Senju, A., Akechi, H., Tojo, Y., Osanai, H., & Hasegawa, T. (2011). Atypical disengagement from faces and its modulation by the control of eye fixation in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *41*, 629–645.

Klein, C., & Fischer, B. (2005). Instrumental and test-retest reliability of saccadic measures. *Biological Psychology*, *68*, 201–213.

Klein, C., & Foerster, F. (2001). Development of prosaccade and antisaccade task performance in participants aged 6 to 26 years. *Psychophysiology*, *38*, 179–189.

Matsuzawa, M., & Shimojo, S. (1997). Infants' fast saccades in the gap paradigm and development of visual attention. *Infant Behavior and Development*, *20*, 449–455.

McConnell, B. A., & Bryson, S. E. (2005). Visual attention and temperament: Developmental data from the first 6 months of life. *Infant Behavior and Development*, *28*, 537–544.

Mosconi, M. W., Kay, M., D'Cruz, A. M., Seidenfeld, A., Guter, S., Stanford, L. D., & Sweeney, J. A. (2009). Impaired inhibitory control is associated with higher-order repetitive behaviors in autism spectrum disorders. *Psychological Medicine*, *39*, 1559–1566.

Nakagawa, A., & Sukigara, M. (2013). Individual differences in disengagement of fixation and temperament: Longitudinal research on toddlers. *Infant Behavior and Development*, *36*(4), 728–735.

Reuter-Lorenz, P. A., Hughes, H. C., & Fendrich, R. (1991). The reduction of saccadic latency by prior offset of the fixation point: An analysis of the gap effect. *Perception & Psychophysics*, *49*, 167–175.

Wass, S. V., Porayska-Pomsta, K., & Johnson, M. H. (2011). Training attentional control in infancy. *Current Biology*, *21*, 1543–1547.

Wass, S. V., & Smith, T. J. (2014). Individual differences in infant oculomotor behavior during the viewing of complex naturalistic scenes. *Infancy*, *19*, 352–384.

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength & Conditioning Research*, *19*(1), 231–240.