Proceedings of Contemporary Ergonomics and Human Factors 2017, pp. 257-265
Available online http://www.ergonomics.org.uk/wp-content/uploads/2015/05/Contemporary-EHF2017-FINAL.pdf

Blink counts can differentiate between task type and load

Rebecca CHARLES and Jim NIXON

Cranfield University, UK

Abstract. Physiological measures have increased in popularity due to the growing availability of equipment allowing their measurement in real-time. Eye blinks are an easy measure to collect using video capture. Our findings indicate that blink counts can differentiate between taskloads and task types during a computer based task, and reflect subjective workload ratings. Blink counts were significantly lower during the tasks involving high visual load when compared to less visually demanding tasks, and lower numbers of blinks were observed under higher taskloads across tasks with a higher visual load. Significant correlations were observed between blink counts and all dimensions of the NASA-TLX for a tracking task, and the mental demand dimension for the combined system monitoring and resource management. No significant correlations were observed for the less visually demanding communications task.

Keywords. Blinks, MATB, taskload, multiple resources

1. Introduction

Physiological measures have been increasingly used in the field of human factors when considering complex cognitive work by using real time information about how a user responds to a task. While this increased use has been documented in many domains (Young, Brookhuis, Wickens, & Hancock, 2014), its growing use is apparent in fixed wing flight operations since the last domain specific review was conducted (Roscoe, 1992). The benefits of being able to monitor the physiological responses of pilots is of growing interest and this work has stemmed from a larger project examining the potential for pilots to extend beyond their performance envelope and for their performance to degrade as a result. This is based on previous work in Air Traffic Control (ATC) (Edwards, Sharples, Wilson, & Kirwan, 2012). This degradation in performance could be due to a number of factors including task load, mental workload or divided attention. If the factors leading to performance degradation can be identified, it may lead to the ability to predict when a pilot may be about to reach the limits of their performance envelope. A further extension to this is investigating the use of adaptive automation or information provision to mitigate reduced performance (Bailey et al., 2006). Such interventions have the ability to lower mental workload and increase situation awareness (Haarmann, Boucsein, & Schaefer, 2009). Eye blinks have the potential to be able to differentiate between tasks using different modalities which could prove to be beneficial when considering pilot state prediction. This article aims to demonstrate the effectiveness of eye blinks as a measure to differentiate between different task types and loads, and the association of eye blinks with subjective workload measures.

1.1 Ocular measures

There have been a variety of ways in which mental workload has been characterised using ocular measures, in experimental (Ryu & Myung, 2005), applied (Hankins & Wilson, 1998) and simulated situations (Veltman & Gaillard, 1996). Tasks involving a high level of visual demand have been shown to yield lower blink rates (Brookings, Wilson, & Swain, 1996; Stern, 1980; Veltman & Gaillard, 1996; Wilson, 2002) than those requiring minimal visual input, for example during auditory tasks (Sirevaag et al., 1993). Pupil diameter has also shown increased changes during tasks involving planning and visual demand when compared to tasks requiring verbal working memory (Causse, Senard, Demonet, & Josette, 2010). In addition, there is evidence to suggest that ocular measures can be used to discriminate between high and low task load, with blink rate decreasing for a high complexity task in a nuclear control context (Hwang et al., 2008), a multi attribute task (Fairclough & Venables, 2006) and general aviation (Hankins & Wilson, 1998; Veltman & Gaillard, 1996). However, Gao et al., (2013) observed latency with blink rate and complexity of a task, and during a tracking task, blink rate was found to decrease with increased complexity (of the tracking task), but not with the increase of complexity of a secondary mental arithmetic task (Ryu & Myung, 2005).

1.2 The MATB

The Multi Attribute Task Battery (MATB) (Comstock & Arnegard, 1992) was developed to provide an experimental platform with which to administer aviation based tasks to non-aviation participants across a range of tasks. The use of the MATB II in experimental studies is documented in the literature, specifically for mental workload studies (Fairclough, Venables, & Tattersall, 2005; Fournier, Wilson, & Swain, 1999; Hsu, Wang, & Chen, 2015; Miyake et al., 2009; Nygren, 1997). The MATB II has demonstrated content validity, construct validity and face validity. The MATB II (figure 1) consists of four tasks: tracking, system monitoring, communications and resource management.

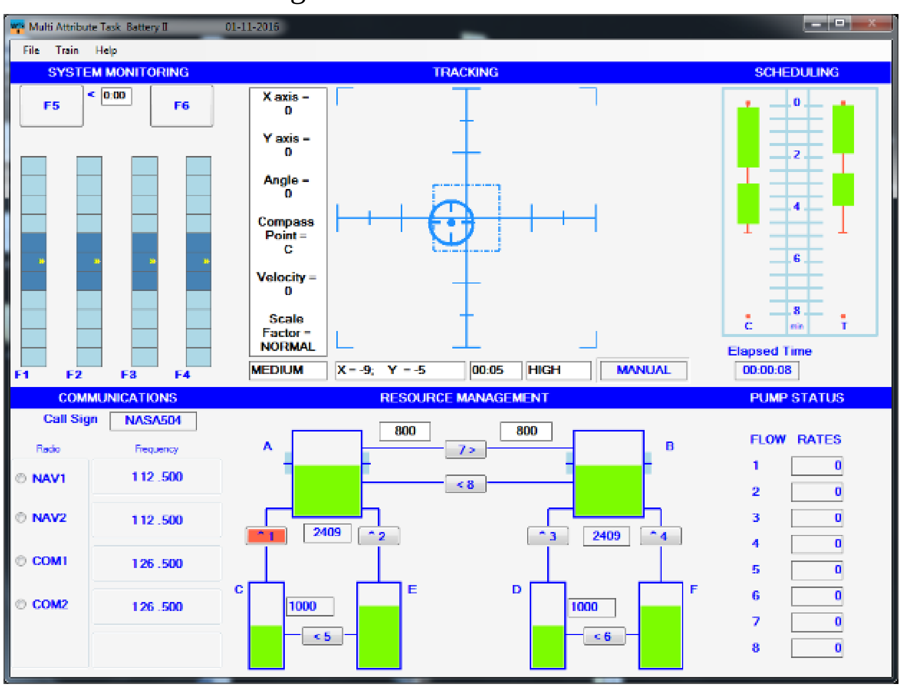

Figure 1 – Screenshot of the MATB II interface (Santiago-Espada, Myer, Latorella, & Comstock, James R., 2011)

The tasks can be classified as different in terms of modalities, codes and outputs using the multiple resources theory (Wickens, 1984, 2008) shown in figure 2a. All of the tasks contain a level of visual content, and require the participant to interact manually with the interface to varying degrees (see figure 2b). The tracking task can be classified as a psychomotor task and requires continuous manual output using the visual-spatial resource. The participant has to constantly respond manually to a visual stimulus ensuring the task is executed according to pre-defined guidelines (that the target remains inside the central box) using a joystick. Due to the nature of the tracking task, the level of visual demand does not change with the taskload (high or low) it is high at both levels. However, the participant may feel that the task is more physically demanding at the high taskload since there is increased resistance when trying to move the target. The system monitoring and resource management tasks require manual output also in the visual spatial resource in response to visual stimuli. During the experiment these two tasks were combined due to their reactive nature. Both of the tasks require the participant to react when certain parameters are reached or exceeded by interacting with the interface. The system monitoring task requires the participant to react by clicking the relevant gauge when it deviates from the central position. The task also requires a response if the button marked F5 lights red, or when the button marked F6, which is normally green, switches off. The resource management task requires the participant to maintain the levels in tanks A and B as close to 2500 as possible, by turning pumps on and off, working around any pump failures. The amount of interaction varies with the low or high taskload, as the number of gauge deviations, light responses and pump failures is manipulated. Unlike the tracking task however, this task requires increased cognitive processing in the form of increased awareness and forward planning. The communications task requires manual output in the visual auditory verbal domain, responding to auditory stimuli. The auditory stimuli either requires a response if it is the correct callsign, or it does not. The taskload levels, high and low for each of the tasks of the MATB-II was determined through a literature search, to establish validated taskload levels.

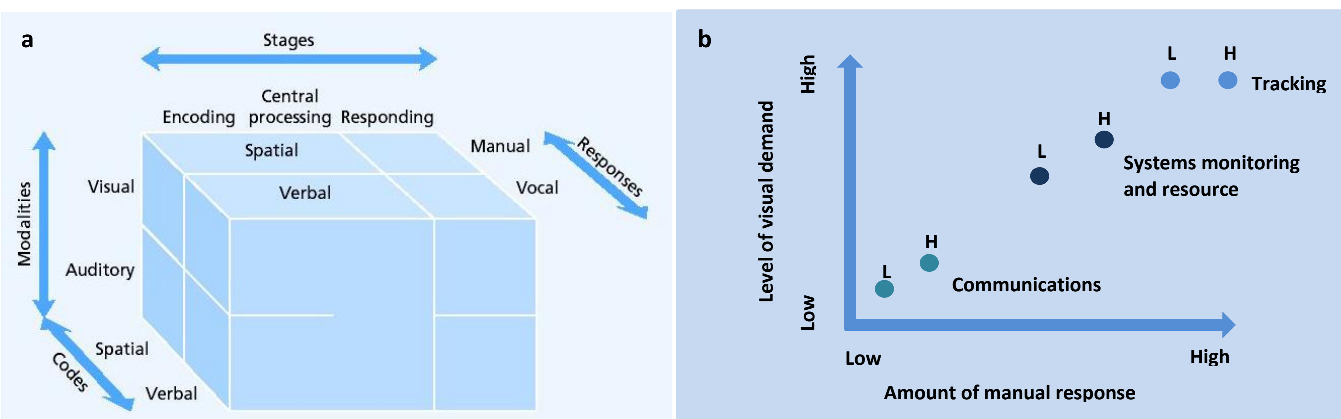

Figure 2 – a) Multiple-resource theory (Wickens, 1984); b) The MATB II tasks (Low (L) and High (H) taskload) plotted according to visual demand and manual response

The NASA-TLX (Hart & Staveland, 1988) was administered after each task, so data was gathered for both the high, and low taskload condition. Individual dimensions are analysed to allow greater diagnosticity.

1.3 Scope

This paper describes the use of eye blinks as a determinate of task type and task load during a computer based task.

Hypothesis 1: Different tasks will elicit different blink counts

Hypothesis 2: Different taskloads will elicit different blink counts

Hypothesis 3: Higher subjective ratings will be associated with lower blink counts

2. Methods

2.1 Participants

Forty four male participants took part in this study, but due to missing data 38 participants were used for this analysis with a mean age of 34.1 years (SD 10.94). All of the participants had normal or corrected to normal vision and none reported consuming alcohol since waking prior to taking part in the experiment. No participants were excluded for health or medication reasons. Four participants also stated that they had flying experience. However, these participants were not professional pilots.

2.2 Design

The Multi Attribute Task Battery 2 (MATB II) was used to deliver tasks to participants. A mixed design was used. The within-subjects factor is task type and taskload. Each participant completed six, five-minute blocks of trials. Three sets of tasks were completed by participants at high and low task loads. These were: resource management and system monitoring, communications, and tracking. Task order was randomized, as was taskload presentation (low to high, or high to low). Three cards with each task detailed on each card were presented face down to the participant. They selected the cards randomly to determine the task order. In addition, 44 pieces of paper were prepared at the start of the experiment, 22 stating the order low to high, and 22 stating the order high to low. Each participant selected a piece of folded paper to determine taskload presentation. Each piece of paper was discarded after each participant. Tasks were presented to participants using the MATB II software provided by NASA. During each block participant faces were filmed using a webcam. The eye blinks were counted manually from the webcam video data post hoc. In addition subjective workload measures using the NASA-TLX were collected.

3. Results

The mean number of blinks observed during the communications task for both the high, and low taskload conditions were considerably higher than for the other tasks (see table 1).

Table 1 – Mean (sd) eye blink counts by condition

	Communications Task	Tracking Task	System monitoring and resource management task
Low Taskload	89.1 (46.7)	40.6 (38.0)	29.5 (19.1)
High Taskload	93.1 (42.5)	31.9 (28.3)	26.6 (17.2)

Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of task type, ($\chi^2(2)=7.95$, $\varepsilon=.84$, $p<0.05$). Therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. There was a significant effect of task type on the number of eye blinks ($F(1.67, 61.77) = 109.24$, $p<.001$, $\eta_p^2 = .75$) (figure 3). Contrasts indicated that the number of eye blinks in the combined resource management and system monitoring task, ($F(1,37)=133.94$, $p<0.001$, $\eta_p^2 = .78$) and the tracking task, ($F(1,37)=138.24$, $p<0.001$, $\eta_p^2 = .79$) were significantly lower than for the communications task. The combined resource management and system monitoring task and communications task approached significance, ($F(1,37)=4.12$, $p=0.05$, $\eta_p^2 = .1$). There was a significant interaction between high and low taskload and the tracking and communications task, ($F(1,37)=15.06$, $p<0.001$, $\eta_p^2 = .29$) . Paired samples t tests revealed that the blink counts were significantly different between the high and low taskload conditions for the tracking task ($t(37) = 2.78$, $p < 0.01$) and the combined resource management and system monitoring task ($t(37) = 2.27$, $p < 0.05$). No significant differences were observed for the communications task.

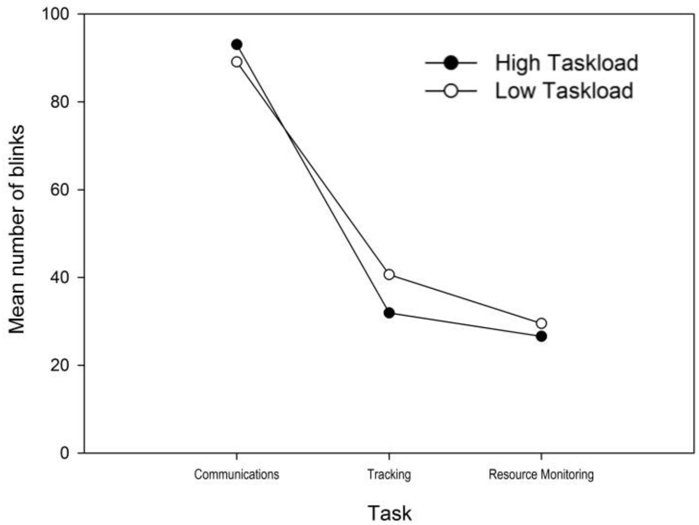

Figure 3 – Eye blink counts by task type and taskload

The NASA-TLX data do not support the use of parametric tests, Spearmans rho is used to assess the direction and magnitude of the associations as a non-parametric alternative in this section. All correlation hypothesis-tests are one-tailed given the directional hypothesis There was a significant negative relationship between the number of eye blinks during the combined resource management and system monitoring task and the NASA-TLX scores for the mental demand dimension, ($r_s = -.22$, N=76, $p < 0.01$), with a higher mental demand relating to a lower number of blinks. For the tracking task significant negative relationships were found for all dimensions of the NASA-TLX; mental demand, ($r_s = -.28$, N=76 $p < 0.01$), effort, ($r_s = -.28$, N=76, $p < 0.01$), temporal demand, ($r_s = -.26$, N=76, $p < 0.05$), frustration, ($r_s = -.29$, N=76 $p < 0.01$), performance, ($r_s = -.31$, N=76 $p < 0.01$), and physical demand, ($r_s = -.30$, N=76), $p < 0.01$). There were no significant relationships found between any of the NASA-TLX scores and the blink counts in the communications task.

4. Discussion and Conclusion

This study addressed three hypotheses:

Hypothesis 1: Different tasks will elicit different blink counts

Hypothesis 2: Different taskloads will elicit different blink counts

Hypothesis 3: Higher subjective ratings will be associated with lower blink counts

Type of task had an effect on the number of eye blinks observed, with the communications task yielding significantly higher blink rates than the visually demanding tasks, supporting hypothesis 1. This finding aligns with previous work (Sirevaag et al., 1993) and demonstrates that there is value in continuing to explore eye blinks as a measure, and predictor of MWL. In addition, there is evidence to suggest that blink rate may only detect short term changes, and level off after time (Fairclough, Venables, & Tattersall, 2005). For subsequent analysis, the blink rates should be considered across all tasks combined rather than each five minute segment individually. In terms of taskload, only the tracking and combined resource management and system monitoring tasks showed significant differences, supporting hypothesis 2. This finding supports previous research, regarding the differences observable in the more visually demanding tasks (Fairclough & Venables, 2006), compared to the less visual, auditory task. This may indicate that blink rate is not a suitable measure of mental workload for tasks consisting primarily of an auditory element. In this context the lack of differences may also indicate that the stimulus for the high and low taskloads for the communications task were not sufficiently different to elicit the changes in workload. This raises an interesting question regarding complexity and how it is defined, and how it is manipulated and implemented in controlled experiments.

The NASA-TLX scores were significantly negatively correlated with the mean number of blinks across all dimensions for the tracking task, and for the mental demand dimension for the system monitoring and resource management task. There were no significant correlations observed for the communications task, so hypothesis 3 is partially supported in that higher subjective ratings for some dimensions are associated with lower blink counts for the more visually demanding tasks. The tracking task requires continuous input and gives continuous feedback; the participant knows instantly if they have made a mistake or the task is not going well. In addition the task becomes more demanding at the higher taskload level as it becomes increasingly difficult to maintain the target in the centre. This may be the explanation for the differences observed across all dimensions (mental demand, effort, temporal demand, frustration, performance, and physical demand) compared to the system monitoring and resource management task in which correlations were observed for the mental demand dimension only. The combined task requires problem solving and forward planning, and although it is possible to establish whether the task is going well or not to a certain degree, there are a number of different items to concentrate on (lights, gauges, fuel levels, and pump failures) which would impact mental demand rather than the tracking task.

The results of this study support existing research suggesting that blink count could be used as an additional measure when investigating mental workload, specifically during visually demanding tasks, and has the ability to differentiate between visual and non-visual task types and taskloads. Our research empirically demonstrates large effect sizes for blink counts associated with high and low visual demand and modest correlations with subjective workload. Another direction would be to consider the blink rate at a more micro level; the recordings suggest that the blink pattern is synched with particular moments of effort, or when mistakes are made, and recognised, by a participant. Interestingly, this effect was apparent in the communications task during the experiment. This may indicate that blink rate could be used to detect when a user believes or recognizes that they have made a mistake, which although not detrimental on its own, could lead to subsequent effects or subjective levels of high workload which may push someone to the bounds of their performance envelope. This may allow the design and implementation of adaptive automation to assist users when they may be struggling. It is also recognized that the MATB II enables tasks in different modalities to be neatly separated and managed, which may not be the case during actual flight. It would be interesting to establish to what extent blink rate can be used during applied environments as an indicator of mental workload when tasks in different modalities are combined.

This article demonstrates that even when using a low technology solution, such as a webcam, eye blinks can successfully be used to differentiate between higher and lower taskloads for visual monitoring tasks., . The next stage in this work is to establish if blink rate is a stable measure of subjective workload and the potential to be a predictor of mental workload which could be harnessed as an input to assess an operators location in a human performance envelope.

Acknowledgements
Many thanks to Emily Carroll for her help with data processing.
This paper is based on work performed in the Programme Future Sky Safety, which has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 640597. The views and opinions expressed in this paper are those of the authors and are not intended to represent the position or opinions of the Future Sky Safety consortium or any of the individual partner organisations.

References

Bailey, N. R., Scerbo, M. W., Freeman, F. G., Mikulka, P. J., & Scott, L. A. (2006). Comparison of a Brain-Based Adaptive System and a Manual Adaptable System for Invoking Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *48*(4), 693–709.

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, *42*(3), 361–377.

Causse, M., Senard, J.-M., Demonet, J. F., & Josette, P. (2010). Monitoring Cognitive and Emotional Processes Through Pupil and Cardiac Responses During Dynamic Versus Logical Task. *Applied Psychophysiology and Biofeedback*, *35*(2), 115–123.

Comstock, J. R., & Arnegard, R. J. (1992). Task Battery Description Monitoring Running the Task Battery.

Edwards, T., Sharples, S., Wilson, J. R., & Kirwan, B. (2012). Factor interaction influences on human performance in air traffic control: the need for a multifactorial model. *Work*, *41 Suppl 1*, 159–66.

Fairclough, S. H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology*, *56*(2), 171–184.

Fairclough, S., & Venables, L. (2006). Prediction of subjective states from psychophysiology: A multivariate approach. *Biological Psychology*.

Fournier, L. R., Wilson, G. F., & Swain, C. R. (1999). Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: Manipulations of task difficulty and training. *International Journal of Psychophysiology*, *31*, 129–145.

Gao, Q., Wang, Y., Song, F., Li, Z., & Dong, X. (2013). Mental workload measurement for emergency operating procedures in digital nuclear power plants. *Ergonomics*, *56*(7), 1070–85.

Haarmann, A., Boucsein, W., & Schaefer, F. (2009). Combining electrodermal responses and cardiovascular measures for probing adaptive automation during simulated flight. *Applied Ergonomics*, *40*(6), 1026–1040.

Hankins, T. C., & Wilson, G. F. (1998). A Comparison of Heart Rate, Eye Activity, EEG and Subjective Measures of Pilot Mental Workload During Flight. *Aviation, Space and Environmental Medicine*, *69*(4), 360–367.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology, 52*, 139-183.

Hsu, B.-W., Wang, M.-J. J., & Chen, C.-Y. (2015). Effective Indices for monitoring mental workload while performing muktiple tasks. *Perceptual and Motor Skills*, *121*(1), 94–117.

Hwang, S.-L., Yau, Y.-J., Lin, Y.-T., Chen, J.-H., Huang, T.-H., Yenn, T.-C., & Hsu, C.-C. (2008). Predicting work performance in nuclear power plants. *Safety Science*, *46*(7), 1115–1124.

Miyake, S., Yamada, S., Shoji, T., Takae, Y., Kuge, N., & Yamamura, T. (2009). Physiological responses to workload change. A test/retest examination. *Applied Ergonomics, 40*(6), 987–996.

Nygren, T. E. (1997). Framing of Task Performance Strategies: Effects on Performance in a Multiattribute Dynamic Decision Making Environment. *Human Factors*, *39*(3), 425–437.

Roscoe, a H. (1992). Assessing pilot workload. Why measure heart rate, HRV and respiration? *Biological Psychology*, *34*(2-3), 259–287.

Ryu, K., & Myung, R. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics*, *35*(11), 991–1009.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–183.

Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock, James R., J. (2011). *The Multi-Attribute Task Battery II (MATB-II) Software for Human Performance and Workload Research: A User's Guide*.

Sirevaag, E. J., Kramer, A. F., Wickens, C. D., Reisweber, M., Strayer, D. L., & Grenell, J. F. (1993). Assessment of pilot performance and mental workload in rotary wing aircraft. *Ergonomics*, *36*(9), 1121–1140.

Stern, J. A. (1980). Aspects of Visual Search Activity Related to Attentional Processes and Skill Development.

Veltman, J. A., & Gaillard, W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological Psychology*, *42*(3), 323–342.

Wickens, C. D. (1984). Processing resources in attention. In Parasuraman, R. & Davies, DR (Eds): *Varieties of Attention*.

Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors*, *50*(3), 449–455.

Wilson, G. F. (2002). An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. *The International Journal of Aviation Psychology*, *12*(1), 3–18.

Young, M. S., Brookhuis, K. a, Wickens, C. D., & Hancock, P. A. (2014). State of science: mental workload in ergonomics. *Ergonomics, 58(1)*(January 2015), 1–17.