Reliability of the 5-min psychomotor vigilance task in a primary school classroom setting

ANDREW WILSON, JAMES DOLLMAN, KURT LUSHINGTON, AND TIMOTHY OLDS
University of South Australia, Adelaide, South Australia, Australia

This study evaluated the reliability of the 5-min psychomotor vigilance task (PVT) in a single-sex Australian primary school. Seventy-five male students (mean age = 11.82 years, $SD = 1.12$) completed two 5-min PVTs using a Palm personal digital assistant (PDA) in (1) an isolated setting and (2) a classroom setting. Of this group of students, a subsample of 37 students completed a test–retest reliability trial within the classroom setting. Using a mixed-model analysis, there was no significant difference in the mean response time (RT) or number of lapses ($RTs \geq 500$ msec) between the isolated and the classroom setting. There was, however, an order effect for the number of lapses in the isolated setting, with the number of lapses being greater if the isolated test was conducted second. Test–retest intraclass correlation coefficients (ICCs) in the classroom setting indicated moderate to high reliability (mean $RT = .84$, lapses = $.59$). Bland–Altman analysis showed no systematic difference between the two settings. Findings suggest that the 5-min PDA PVT is a reliable measure of sustained attention in the classroom setting in this sample of primary-aged schoolchildren. The results provide further evidence for the versatility of this measuring device for larger interventions outside the laboratory.

Sustained attention during lesson time is important for student learning (Hunt & Randhawa, 1983; Muris, 2006; National Institute of Child Health and Human Development, 2003; Zentall, 2005). Accordingly, several groups have developed measures of sustained attention for research purposes that are suitable for classroom-based testing (Conners & MHS Staff, 2000; Leark, Greenberg, Kindschi, Dupuy, & Hughes, 2007). Currently, however, the majority of instruments are time-consuming and/or computer (PC) dependent, and there is a need for objective measures that can be administered easily in the classroom to large cohorts of students with minimal distraction to the learning environment. A promising candidate is the psychomotor vigilance task (PVT; Ambulatory Monitoring, Inc., Ardsley, NY). The PVT requires participants to press a response button when a randomly sequenced target stimulus is presented and provides two measures of sustained attention: response time (RT) and lapses (i.e., $RTs > 500$ msec). The PVT software (Walter Reed Army Institute) can be easily downloaded onto a handheld personal digital assistant (PDA).

The PDA PVT has been used to measure sustained attention in studies of sleep and fatigue in adults (Blatter et al., 2006; Graw, Kräuchi, Knoblauch, Wirz-Justice, & Cajochen, 2004; Heishman, 2007; Loh, Lamond, Dorian, Roach, & Dawson, 2004; Petrilli, Roach, Dawson, & Lamond, 2006) and to a lesser extent in children (Beijamini, Silva, Peixoto, & Louzada, 2008; Peters et al., 2009; Venker et al., 2007), but notably, the studies with children were conducted only in laboratory settings. In these

studies, the children were withdrawn from group settings and isolated at the time of testing (Beijamini et al., 2008; Peters et al., 2009; Venker et al., 2007). Although such testing is appropriate in the laboratory setting, removing children from the classroom disrupts learning and limits data collection. Therefore, before commencing trials with the PDA PVT testing in schools, we felt it would be of benefit to explore whether it is necessary to remove children and test them in isolation or if they can be tested in groups in the classroom. The aim of the present study was to assess performance of the PDA PVT in a classroom in comparison with individual test environments.

METHOD

Sample

Seventy-five male students (mean age = 11.82 years, $SD = 1.12$) were recruited as participants from a local elementary school (two Year 5 classes [10-year-olds] and two Year 7 classes [12-year-olds]). We excluded children with injuries that might affect PDA administration and children with an intellectual disability. Children were asked not to consume caffeinated products on the day of testing.

5-Min PDA PVT

The 5-min PDA PVT detailed the students' sustained attention over a 5-min period. We recorded the mean RT (in milliseconds) to a visual display and the number of lapses ($RTs \geq 500$ msec). Participants were instructed to press a nominated button using their dominant hand in response to a repeating visual stimulus presented randomly at 1- to 5-sec intervals over a 5-min period. The PDA-specified PVT software developed by the Walter Reed Army Institute of Research (Thorne et al., 2005) was ported onto a commer-

A. Wilson, andrew.wilson@postgrads.unisa.edu.au

Table 1
Mean ($\pm$ SD) 5-min PDA PVT Response Times and Lapses
Collected Under Isolated and Classroom (Group) Conditions,
Together With Interclass Correlation Coefficients (ICCs)

PVT Measure	Testing Environment				ICC [†]
	Isolation (<i>n</i> = 75)		Classroom (<i>n</i> = 75)		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Response time (msec)	342.7	74.6	344.1	71.4	.82
Number of lapses ^{††}	6.9	6.8	6.5	5.7	.67

[†]Mean isolation vs. mean classroom. ^{††}A lapse is recorded as any response $\geq$ 500 msec.

cially available M105 Palm Pilot device (Palm Inc., Sunnyvale, CA), a small compact device ($7.4 \times 1.8 \times 11.8$ cm, 137 g) with an LCD screen. The PDA was configured so that remaining PDA functions were switched off and the left or right response button was selected according to hand preference.

The same researcher instructed all students about how to complete the 5-min PDA PVT using the same protocol for each group of students. To assess for practice effects, as reported by Dinges et al. (1997) using adult participants, three practice trials were conducted using students (*n* = 61) with a mean age of 12.08 years (*SD* = 1.52). The mean RT results for the three trials were 390.99, 334.0, and 334.97 msec and the numbers of lapses were 8.1, 5.2, and 5.9. There was a change in values for both lapses and mean RT from the first trial to the third, indicating that a practice effect is evident. Although no significant difference existed between the second and third trials, we decided to adopt three practice trials, as suggested in previous studies (Baulk, Biggs, Reid, van den Heuvel, & Dawson, 2008; Dinges et al., 1997; Lamond, Petrilli, Dawson, & Roach, 2006). The participants from the main study conducted their three practice trials in the week prior to their trials in the isolated and the classroom environments.

Setting and Testing Protocol

We used a crossover repeated measures design in this study. Students were tested in year levels and were randomly assigned to two groups. Group 1 started with a 5-min PDA PVT test in a classroom setting followed by a 5-min test in an isolated setting immediately thereafter, whereas the order was reversed for Group 2. The year levels were tested on different days, 1 day apart, with testing conducted in the morning for both year levels.

The setting for the classroom condition was a standard classroom with visual displays on the walls, desks in rows, and so on. In order to simulate the protocols used in previous PVT studies (Lamond, Dawson, & Roach, 2005; Peters et al., 2009), we set up the testing for the isolated condition in a standard classroom but with blank walls, with participants seated facing a blank wall in their own bay isolated by two side partitions (1.6×1.2 m).

The students were instructed to start the test at the same time in each setting and not to talk during the testing. To ensure that feed-

back did not act as a motivating factor, or result in students calling out their scores during the test, the feedback component of the 5-min PDA PVT was disabled.

Reliability

To determine the test-retest reliability of the 5-min PDA PVT in the classroom condition, we had 37 participants in the main study complete an additional 5-min PDA PVT within 24 h in the same classroom.

Statistical Analysis

The standard PalmPVT 2.0.1 Windows Companion Program (PC software) that comes with the PVT installation program was used to generate the mean RT (inclusive of lapses) and the number of lapses.

We used a mixed-model design in the main study to compare mean RT with number of lapses in relation to any testing environment effects or for any order effect. For the test-retest reliability study, we used Bland-Altman analysis to assess the level of agreement between the repeated test scores. To further determine reliability, we also generated a single measure intraclass correlation coefficient (ICC; i.e., pooled mean [*SD*] of RT and the number of lapses) between the two testing environments. We summarized descriptive statistics using untransformed data. Data used in further analysis were checked for normality. To normalize data, we log-transformed the number of lapses and generated a reciprocal value for the mean RT value ($1/\text{MRT} \times 1,000$). The statistical software packages SPSS 15.0 for Windows (SPSS, Inc., Chicago, IL) and MedCalc for Windows 10.4.8.0 (MedCalc Software, Mariakerke, Belgium) were used for analysis of data. Outliers were identified as any value greater than 3 *SDs* from the mean. Values identified in the descriptive statistics (see Table 1) are shown with the outliers removed. All of the data we analyzed excluded outliers; in total, one outlier was removed from the data set (lapse value in isolation setting).

Approval for this study was granted by the University of South Australia's Human Research Ethics Committee and the school's governing educational office. Parents and participants gave informed consent to participate in the study.

RESULTS

Preliminary analysis revealed no significant difference between the mean RT in the isolated and classroom environments (see Table 1). The number of lapses, however, was slightly higher in the isolated than in the classroom environment (see Table 1).

Mixed-model analysis indicated that performance in each of the two 5-min PDA PVT metrics (mean RT and lapses) was not significantly affected by the testing environment (Table 2).

For mean RT, no administration order effect was evident (Table 3). The results for the number of lapses, how-

Table 2
Linear Mixed-Model Analysis of Test Setting

	Classroom vs. Isolation		95% Confidence Interval for Difference	
	Mean Difference*	Significance	Lower Bound	Upper Bound
Mean response time (msec)	1.3	.72 ^a	-10.0	12.7
Number of lapses [†]	-0.58	.64 ^b	-1.7	0.6

*Difference calculated by classroom mean minus isolation mean. ^aReciprocal of the mean RT. ^bLog-transformed. [†]A lapse is recorded as any response $\geq$ 500 msec.

Table 3
Linear Mixed-Model Analysis of the Order of Testing

	Order of Testing		95% Confidence Interval for Difference	
	Mean Difference*	Significance	Lower Bound	Upper Bound
Mean response time (msec)	0.74	.82 ^a	-10.6	12.1
Number of lapses [†]	0.92	.01 ^b	-0.2	2.0

*Difference calculated by 2nd trial minus 1st trial. ^aReciprocal of the mean RT. ^bLog-transformed. [†]A lapse is recorded as any response ≥ 500 msec.

ever, indicated that the order of testing was a predictor of this measure. In the second test, the number of lapses was significantly higher than in the first ($p = .01$).

Among the subsample of students who completed the 5-min PVT twice (test–retest) within the classroom environment (Table 4), values were marginally higher in the first test for mean RT and marginally lower for the number of lapses.

The intraclass correlations were moderate for both mean RT (.84) and the number of lapses (.60). Bland–Altman analysis showed that the biases were small and not significantly different from zero (Figures 1 and 2).

DISCUSSION

In the past, the PVT-192 (Ambulatory Monitoring, Inc.) has been extensively used as an independent measure of mental fatigue or attention levels based on RT measures (Tyagi, Shen, Shao, & Li, 2009), but due to its size and cost per unit, it has been best suited to laboratory-style testing. The development by the Walter Reed Army Institute of Research of the 5-min PVT, designed to be run on Palm operating systems, had the potential to offer researchers two major improvements: The first was reduced testing time, while maintaining the sensitivity of the device to measuring the vigilance of participants (Lamond et al., 2005; Lamond et al., 2006; Roach, Dawson, & Lamond, 2006). Second, the reduced cost of the devices made larger sample sizes possible.

For researchers interested in the attention levels of students, this combination of low cost and shorter test times meant that the 5-min PDA PVT was more suited to school-based interventions than previous PVT devices. Yet, to date, researchers have been slow to adopt this new version of the PVT as a means of objective measurement in school interventions. Current interventions looking into ways to improve the sustained attention of students

in schools have been largely confined to small studies of students with attention deficit hyperactivity disorder (ADHD).

Measurement of sustained attention is usually conducted in isolation, to minimize external distractions. For large, school-based interventions, the disruption and consequent loss of valuable learning time as a result of having small groups tested away from the classroom is a potential concern for school administrators. The results of this study offer researchers and schools an alternative, which does away with the need to test in isolation, in small groups, and over long periods of time. For this sample, there was no significant difference in the measures recorded by the 5-min PDA PVT when it was conducted in the students' own classroom, in comparison with the smaller, more traditional, isolated environment. The mean RT in both settings showed a high level of test–retest reliability. There were also no differences in the numbers of lapses between the two settings.

The order effect identified during analysis appears to be confined to the isolated setting. The higher number of lapses in the isolated setting when participants completed the second 5-min PVT appears to suggest a decline in sustained attention; however, no such decline is evident in the mean RT measure or between lapses in the two classroom tests. Whether the additional stimuli present in the classroom result in a higher level of sustained attention is unknown. Alternatively, it may be that the quiet, isolated environment is unmasking fatigue in the students, suggesting a possible need to measure participants' sleep patterns in future studies.

A limiting factor of this study is that it was completed in an all-male primary school, so generalization of the results beyond these circumstances should be done with caution. Previous studies (Bejjamini et al., 2008) have reported differences in mean RT between the sexes and in their approach to the task, particularly regarding the

Table 4
Reliability (Test–Retest) Statistics

	Classroom Environment				Pearson's Correlation
	Test 1		Test 2		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
<i>n</i> = 36					
Response time (msec)	307.9	50.0	302.8	45.8	.84 (<.0001)
Number of lapses [†]	3.6	3.1	3.7	4.0	.62 (.0002)

[†]A lapse is recorded as any response ≥ 500 msec.

Figure 1. Bland–Altman analysis: Correlation of mean response time.

Figure 2. Bland–Altman analysis: Correlation of the number of lapses.

accuracy of the response (Blatter et al., 2006). Therefore, future research should consider expanding the sample to include females and secondary students.

The results of this study indicate that the 5-min PDA PVT has the potential to be an effective measuring tool of sustained attention in school-based interventions due to its portability, short duration, and minimal learning curve (Dinges et al., 1997). Pilot data ($n = 61$) conducted prior to the present investigation suggested that second and third practice trials of the 5-min PVT in students indicated no significant improvements or changes in measurements for RTs and lapses. This finding is smaller than the reported learning curve for adult trials (Dinges et al., 1997), which may suggest that the quantity of practice trials in students or children need not be as high as that for adults.

The low number of practice trials, combined with the results from the main study, add further evidence to the device's potential; results suggest that for this particular sample, time-consuming testing in isolation is not necessary. The implication of this finding is a potential reduction in the impact of PVT assessment on the classroom learning environment, which could result in more interest from educational institutions and staff members in being involved in future studies.

Conclusion

Results from this study give further evidence about the potential of the 5-min PDA PVT to be an effective measuring tool of sustained attention in school-based interventions. The results suggest that for this particular sample, isolated testing environments are not necessary. The

implication of this finding is a potential reduction in the disruption that interventions have on the classroom learning environment, paving the way for potentially greater numbers of school-based studies.

Practical Implications

The findings of this study suggest that for this sample, it is possible to conduct the 5-min PDA PVT in a normal classroom setting. This opens up the possibility for more widespread use of the PVT in schools or similar environments. Previously, the need for testing in isolation had made studies in such environments impractical. This should give researchers the potential for greater access to more objective quantitative data on students' sustained attention.

AUTHOR NOTE

This study was conducted without financial assistance from external funding sources. Correspondence concerning this article may be addressed to A. Wilson, Division of Health Sciences, University of South Australia, City East Campus Rm. P7-27, GPO Box 2471, Adelaide, South Australia 5001, Australia (e-mail: andrew.wilson@postgrads.unisa.edu.au).

REFERENCES

- BAULK, S. D., BIGGS, S. N., REID, K. J., VAN DEN HEUVEL, C. J., & DAWSON, D. (2008). Chasing the silver bullet: Measuring driver fatigue using simple and complex tasks. *Accident Analysis & Prevention*, **40**, 396-402. doi:10.1016/j.aap.2007.07.008
- BEJAMINI, F., SILVA, A. G. T., PEIXOTO, C. A. T., & LOUZADA, F. M. (2008). Influence of gender on psychomotor vigilance task performance by adolescents. *Brazilian Journal of Medical & Biological Research*, **41**, 734-738. doi:10.1590/s0100-879x2008000800016
- BLATTER, K., GRAW, P., MÜNCH, M., KNOBLAUCH, V., WIRZ-JUSTICE, A., & CAJOCHEN, C. (2006). Gender and age differences in psychomotor vigilance performance under differential sleep pressure conditions. *Behavioural Brain Research*, **168**, 312-317. doi:10.1016/j.bbr.2005.11.018
- CONNERS, C. K., & THE MHS STAFF (EDS.) (2000). *Conners' Continuous Performance Test II: Computer program for Windows technical guide and software manual*. North Tonawanda, NY: Multi-Health Systems.
- DINGES, D. F., PACK, F., WILLIAMS, K., GILLEN, K. A., POWELL, J. W., OTT, G. E., ET AL. (1997). Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. *Sleep*, **20**, 267-277.
- GRAW, P., KRÄUCHI, K., KNOBLAUCH, V., WIRZ-JUSTICE, A., & CAJOCHEN, C. (2004). Circadian and wake-dependent modulation of fastest and slowest reaction times during the psychomotor vigilance task. *Physiology & Behavior*, **80**, 695-701. doi:10.1016/j.physbeh.2003.12.004
- HEISHMAN, S. J. (2007). Pharmacodynamics. In S. B. Karch (Ed.), *Drug abuse handbook* (2nd ed., pp. 207-312). Boca Raton, FL: CRC Press.
- HUNT, D., & RANDHAWA, S. B. (1983). Cognitive processes and achievement. *Alberta Journal of Educational Research*, **29**, 206-215.
- LAMOND, N., DAWSON, D., & ROACH, G. D. (2005). Fatigue assessment in the field: Validation of a hand-held electronic psychomotor vigilance task. *Aviation, Space, & Environmental Medicine*, **76**, 486-489.
- LAMOND, N., PETRILLI, R. M., DAWSON, D., & ROACH, G. D. (2006). Do shorter international layovers allow sufficient opportunity for pilots to recover? *Chronobiology International*, **23**, 1285-1294. doi:10.1080/07420520601062387
- LEARK, R. A., GREENBERG, L. K., KINDSCHI, C. L., DUPUY, T. R., & HUGHES, S. J. (2007). *Test of Variables of Attention: Professional manual*. Los Alamitos, CA: TOVA Co.
- LOH, S., LAMOND, N., DORRIAN, J., ROACH, G., & DAWSON, D. (2004). The validity of psychomotor vigilance tasks of less than 10-minute duration. *Behavior Research Methods, Instruments, & Computers*, **36**, 339-346.
- MURIS, P. (2006). Relation of attention control and school performance in normal children. *Perceptual & Motor Skills*, **102**, 78-80. doi:10.2466/PMS.102.1.78-80
- NATIONAL INSTITUTE OF CHILD HEALTH AND HUMAN DEVELOPMENT, EARLY CHILD CARE RESEARCH NETWORK (2003). Do children's attention processes mediate the link between family predictors and school readiness? *Developmental Psychology*, **39**, 581-593. doi:10.1037/0012-1649.39.3.581
- PETERS, J. D., BIGGS, S. N., BAUER, K. M. M., LUSHINGTON, K., KENNEDY, D., MARTIN, J., & DORRIAN, J. (2009). The sensitivity of a PDA-based psychomotor vigilance task to sleep restriction in 10-year-old girls. *Journal of Sleep Research*, **18**, 173-177. doi:10.1111/j.1365-2869.2008.00716.x
- PETRILLI, R. M., ROACH, G. D., DAWSON, D., & LAMOND, N. (2006). The sleep, subjective fatigue, and sustained attention of commercial airline pilots during an international pattern. *Chronobiology International*, **23**, 1347-1362. doi:10.1080/07420520601085925
- ROACH, G. D., DAWSON, D., & LAMOND, N. (2006). Can a shorter psychomotor vigilance task be used as a reasonable substitute for the ten-minute psychomotor vigilance task? *Chronobiology International*, **23**, 1379-1387. doi:10.1080/07420520601067931
- THORNE, D. R., JOHNSON, D. E., REDMOND, D. P., SING, H. C., BELENKY, G., & SHAPIRO, J. M. (2005). The Walter Reed palm-held psychomotor vigilance test. *Behavior Research Methods*, **37**, 111-118.
- TYAGI, R., SHEN, K., SHAO, S., & LI, X. (2009). A novel auditory working-memory vigilance task for mental fatigue assessment. *Safety Science*, **47**, 967-972. doi:10.1016/j.ssci.2008.10.018
- VENKER, C. C., GOODWIN, J. L., ROE, D. J., KAEMINGK, K. L., MULVANEY, S., & QUAN, S. F. (2007). Normative psychomotor vigilance task performance in children ages 6 to 11: The Tucson Children's Assessment of Sleep Apnea (TuCASA). *Sleep Breath*, **11**, 217-224. doi:10.1007/s11325-007-0103-4
- ZENTALL, S. S. (2005). Theory- and evidence-based strategies for children with attentional problems. *Psychology in the Schools*, **42**, 821-836. doi:10.1002/pits.20114

(Manuscript received December 17, 2009;
accepted for publication February 16, 2010.)