Test-Retest Reliability of Virtual Reality Measurements of Perceptual-Motor Function among Healthy College Students

THE UNIVERSITY of TENNESSEE CHATTANOOGA

Noblett CN, Sarris CE, Wilkerson GB

Background

- Response time (RT) typically assessed with computerized neurocognitive tests[1]
  - Whole-body RT may better reflect sport performance capability than finger presses
- Brain processing efficiency can be impaired by a neuroinflammatory process[2]
  - Optimal perceptual-motor processing essential for injury avoidance
- Current clinical methods inadequate to detect subtle brain processing impairment[3]
  - Virtual reality (VR) offers a method to precisely measure stimulus-response time

Background

- Time between stimulus and initial response estimates brain processing speed
- Most research has assessed central tendency of multi-trial performance (mean)
- Intra-individual variability (trial-to-trial inconsistency) reflects brain efficiency[4]
- Test-retest reliability of VR metrics have not been documented[5]
  - Study purpose: To assess the consistency of mean and intra-individual variability values acquired from an immersive VR system on 3 successive days

Participants

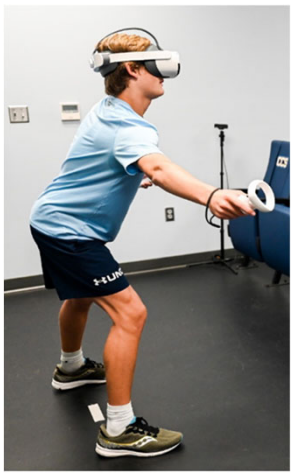
- 24 healthy graduate students volunteered to participate
- Analysis limited to 19 participants with valid and complete data (9 males and 10 females)
  - ≥ 75% response accuracy for 40-trial test
  - Data acquired for all 3 test sessions

Male:	Age: 24.0 ± 1.4 (Range: 22 - 26)
	Height: 177.2 ± 5.0 cm
	Weight: 80.0 ± 14.2 kg
Female:	Age: 23.0 ± 1.6 (Range: 21 - 27)
	Height: 163.6 ± 9.1 cm
	Weight: 74.2 ± 20.4 kg

#22-117

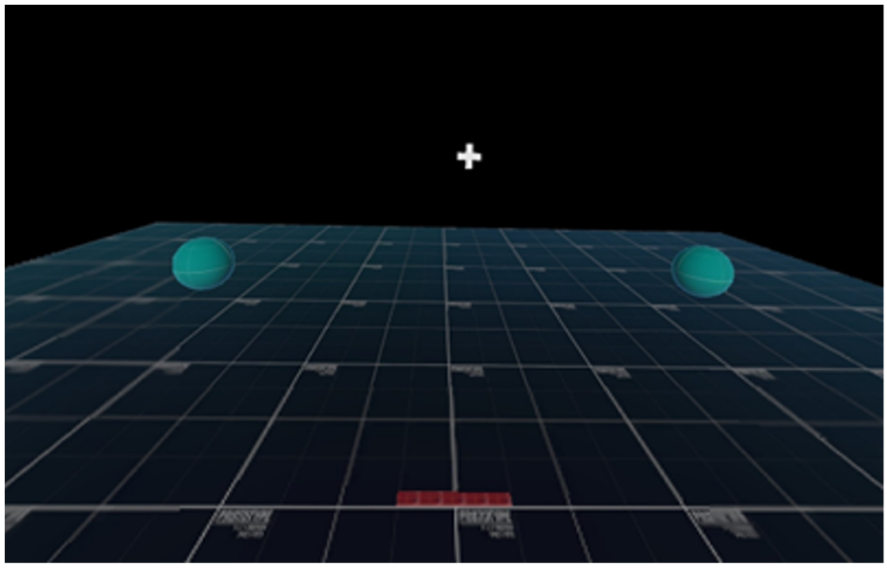
Methods: VR Test Procedure

- 40 successive reaching/lunging responses to visual stimuli

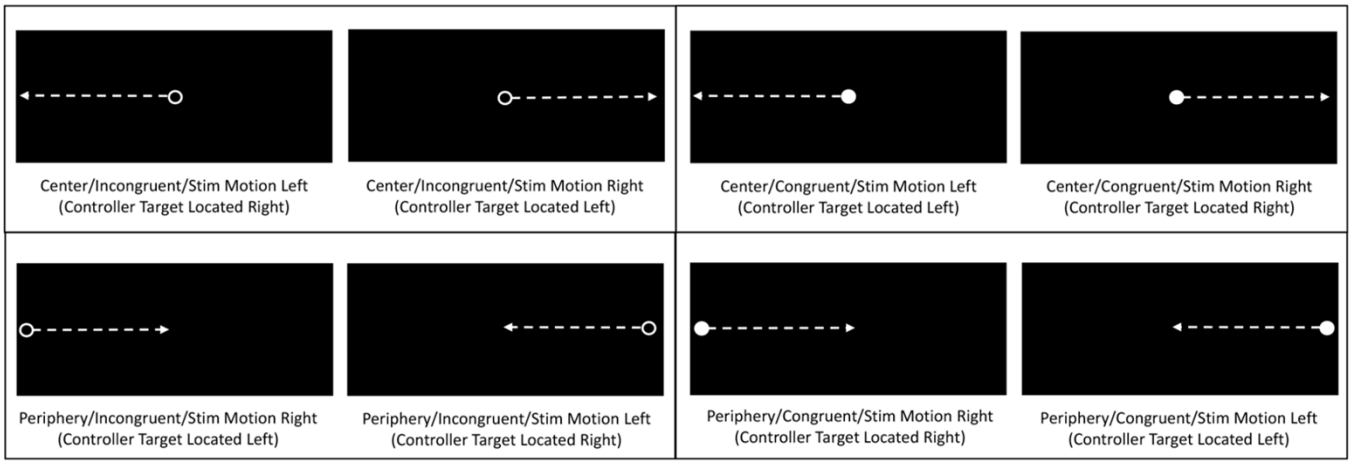


  - Left vs. Right response determined by visual stimulus characteristics
  - Reaching/lunging distance based on T-pose measurement (80%)
  - Response targets located beyond peripheral margin of visual field
  - VR hand controller used to register response to visual stimulus
  - Auditory and haptic feedback confirmed contact with response target

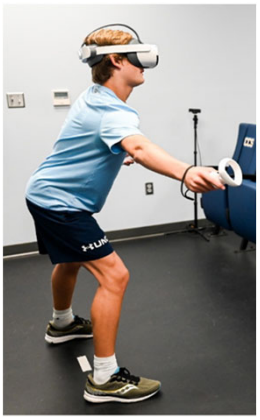
Methods: Headset View Prior to Start of VR Test

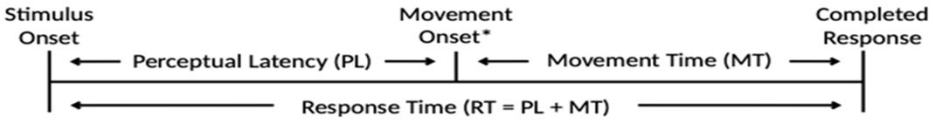


Methods: 8 Different Trial Types



Operational Definitions for Time Intervals





* 6° Angular Rotation (Eyes and Neck) or 10 cm Linear Translation (Arm and Step)

Operational Definitions of Perceptual Latency and Response Time

- Perceptual Latency: Eyes (6˚), Neck (6˚), Arm (10 cm), Step (10 cm)
- Response Time: Eyes (Max), Neck (Max), Arm (Max), Step (Max)

Methods: VR Test Metrics

- Perceptual Latency (PL): Eyes, Neck, Arm, Step
- Response Time (RT): Eyes, Neck, Arm, Step
- Composite Metric: Rate Correct Score (RCS)
  - Calculated from Arm Movements (Hand Controller)
    - RCS-PL = Number Correct / Sum of PL Values
    - RCS-RT = Number Correct / Sum of RT Values

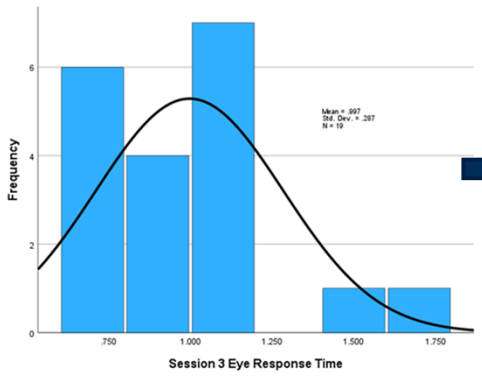
Test-Retest Reliability (Consistency of Repeated Measures)

- Assessment of distribution normality
  - Skew value (< 0 or > 0)
  - Shapiro-Wilk test (P < .05 indication of deviation from normality)
    - Natural log (Log_e) transformation of positively skewed data
- Intra-Class Correlation Coefficient (ICC)
  - Interpretation[6]
    - Poor: < .500
    - Moderate: .500 to .749
    - Good: .750 to .890
    - Excellent: ≥ .900

Statistical Analysis

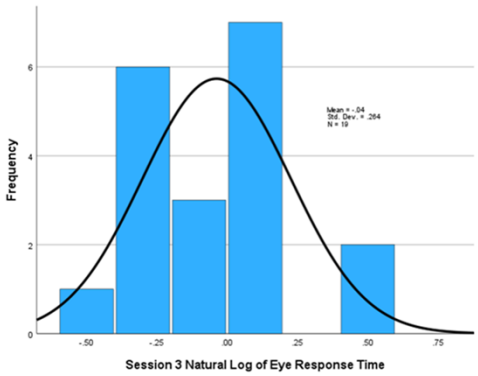
- Assessment of distribution normality
- Natural log (Log_e) transformation of positively skewed distribution

Original Eye Response Time Data



Skew = 1.311 Shapiro-Wilk P = .019

Transformed Eye Response Time Data (Natural Log)



Skew = 0.648 Shapiro-Wilk P = .293

11

Table 1. Distribution Skew and Shapiro-Wilk Test Result ($P_{S\text{-}W}$) for 40-Trial Mean of Original Data and Natural Log (Log_e) Transformation; Test-Retest at 24-hour Intervals; n=19 Graduate Students

	Session 1				Session 2				Session 3			
Metric	Skew	(Log_e)	$P_{S\text{-}W}$	(Log_e)	Skew	(Log_e)	$P_{S\text{-}W}$	(Log_e)	Skew	(Log_e)	$P_{S\text{-}W}$	(Log_e)
Eye Perceptual Latency	2.283	(1.528)	<.001	(.017)	0.784	(0.105)	.265	(.683)	0.744	(0.208)	.199	(.618)
Neck Perceptual Latency	2.412	(1.666)	<.001	(.012)	1.425	(0.688)	.038	(.533)	0.899	(0.420)	.072	(.271)
Arm Perceptual Latency	1.117	(0.626)	.111	(.590)	0.541	(0.109)	.726	(.969)	-0.042	*	.847	*
Step Perceptual Latency	2.196	(1.443)	.001	(.041)	1.060	(0.427)	.038	(.209)	-0.030	*	.804	*
Rate Correct Score - PL	-0.094	*	.810	*	0.313	(-0.071)	.930	(.983)	0.387	(-0.072)	.984	(1.000)
Eye Response Time	-0.306	*	.564	*	0.781	(-0.066)	.535	(.988)	1.311	(0.648)	.019	(.293)
Neck Response Time	2.647	(1.817)	<.001	(.006)	0.349	(-0.033)	.965	(.995)	-0.125	*	.760	*
Arm Response Time	2.264	(1.498)	<.001	(.034)	0.479	(0.226)	.618	(.811)	0.444	(0.310)	.053	(.058)
Step Response Time	2.543	(1.747)	<.001	(.006)	0.320	(0.102)	.381	(.516)	0.110	(0.024)	.058	(.056)
Rate Correct Score - RT	-0.569	*	.483	*	-0.029	*	.880	*	0.242	0.069	.259	(.347)

* Log_e transformation increases negative skew

12

Table 2. Distribution Skew and Shapiro-Wilk Test Result (P_{S-W}) for 40-Trial Intra-Individual Variability of Original Data and Natural Log (Log_e) Transformation; Test-Retest at 24-hour Intervals; n=19 Graduate Students

	Session 1				Session 2				Session 3			
Metric	Skew	(Log_e)	P_{S-W}	(Log_e)	Skew	(Log_e)	P_{S-W}	(Log_e)	Skew	(Log_e)	P_{S-W}	(Log_e)
Eye Perceptual Latency	0.964	(0.357)	.033	(.285)	0.754	(0.054)	.299	(.827)	1.669	(0.255)	.010	(.790)
Neck Perceptual Latency	1.563	(0.585)	.002	(.216)	1.486	(0.323)	.010	(.857)	1.969	(0.524)	.003	(.945)
Arm Perceptual Latency	3.679	(1.230)	<.001	(.065)	1.211	(0.154)	.200	(.796)	3.873	(1.600)	<.001	(.012)
Step Perceptual Latency	1.427	(0.119)	.009	(.629)	0.999	(0.125)	.077	(.695)	3.916	(1.807)	<.001	(.005)
Eye Response Time	-0.261	*	.813	*	-0.201	*	.953	*	0.385	(0.007)	.627	(.812)
Neck Response Time	1.639	(0.779)	.002	(.165)	1.427	(0.393)	.013	(.738)	0.867	(0.098)	.123	(.933)
Arm Response Time	2.297	(0.845)	<.001	(.384)	1.880	(0.660)	.003	(.515)	0.468	(0.078)	.092	(.286)
Step Response Time	2.616	(0.913)	<.001	(.210)	1.798	(0.694)	.003	(.282)	4.027	(2.259)	<.001	(<.001)

* Log_e transformation increases negative skew

13

Natural Log Transformation

- 40-Trial Mean Values
  - Normality improved
    - Session 1: 7 of 10 metrics
    - Session 2: 9 of 10 metrics
    - Session 3: 7 of 10 metrics
- 40-Trial Intra-Individual Variability Values
  - Normality improved
    - Session 1: 7 of 8 metrics/sessions
    - Session 2: 7 of 8 metrics/sessions
    - Session 3: 8 of 8 metrics/sessions

14

Table 3. Geometric Mean Value (Natural Log Transformation of Original Mean Value); Test-Retest at 24-hour Intervals; n=19 Graduate Students

	Session 1		Session 2		Session 3		Sessions 1-2-3	
Metric	G Mean	(Log_e)	G Mean	(Log_e)	G Mean	(Log_e)	ICC (2,k)	P_{Diff}
Eye Perceptual Latency	0.558	(-0.584)	0.563	(-0.575)	0.593	(-0.523)	.903	.150
Neck Perception Latency	0.655	(-0.423)	0.629	(-0.463)	0.616	(-0.484)	.922	.130
Arm Perceptual Latency	0.739	(-0.302)	0.680	(-0.385)	0.663	(-0.411)	.884	<.001
Step Perceptual Latency	0.760	(-0.275)	0.708	(-0.346)	0.699	(-0.358)	.907	.004
Eye Response Time	0.826	(-0.191)	0.964	(-0.037)	1.038	(-0.038)	.618	.052
Neck Response Time	1.068	(0.066)	1.017	(0.017)	0.995	(-0.005)	.904	.023
Arm Response Time	1.261	(0.232)	1.201	(0.183)	1.151	(0.141)	.837	.001
Step Response Time	1.307	(0.267)	1.247	(0.221)	1.216	(0.196)	.882	.054

G Mean: Geometric Mean (Estimated Median of Original Data [Back-Transformation of Log_e Value])
ICC (2,k): Two-Way Random Effects, Absolute Agreement, Average of Measures Intraclass Correlation Coefficient
P_{Diff}: P-value for Repeated Measures Analysis of Variance Difference among Sessions

15

Table 4. Pairwise Comparisons of Natural Log Transformation of Original Mean Value; Test-Retest at 24-hour Intervals; n=19 Graduate Students

	Session 1 – Session 2					Session 2 – Session 3				
Metric	ICC (2,k)	P_{Diff}	%Diff	(95% CI)	%CV	ICC (2,k)	P_{Diff}	%Diff	(95% CI)	%CV
Eye Perceptual Latency	.886	.776	-0.8	(-6.6, 5.3)	13.2	.867	.125	-5.4	(-11.3, 1.6)	15.1
Neck Perceptual Latency	.919	.142	4.1	(-1.5, 10.0)	12.1	.934	.330	2.1	(-2.3, 6.5)	9.3
Arm Perceptual Latency	.856	.001	8.7	(3.9, 13.6)	9.7	.937	.093	2.7	(-0.5, 5.9)	6.6
Step Perceptual Latency	.909	<.001	7.4	(3.6, 11.4)	7.9	.925	.525	1.1	(-2.5, 4.9)	7.9
Eye Response Time	.367	.063	-14.3	(-27.2, 0.9)	40.3	.591	.988	0.1	(-12.3, 14.2)	31.5
Neck Response Time	.886	.059	5.0	(-0.2, 10.5)	11.1	.908	.247	2.2	(-1.6, 6.2)	8.3
Arm Response Time	.804	.084	5.0	(-0.7, 11.1)	12.4	.887	.016	4.3	(0.9, 7.9)	7.2
Step Response Time	.790	.114	4.8	(-1.2, 11.1)	13.0	.869	.186	2.5	(-1.3, 6.6)	8.3

ICC (2,k): Two-Way Random Effects, Absolute Agreement, Average of Measures Intraclass Correlation Coefficient
P_{Diff}: P-value for Paired Samples t-Test Difference between Sessions
%Diff: Symmetric Percentage Difference between Sessions (95% Confidence Interval)
%CV: Percentage Coefficient of Variation between Sessions

16

Table 5. Geometric Mean of Intra-Individual Variability (Natural Log Transformation of Original Intra-Individual Variability Value): Test-Retest at 24-hour Intervals; n=19 Graduate Students

	Session 1		Session 2		Session 3		Sessions 1-2-3	
Metric	G Mean	(Log_e)	G Mean	(Log_e)	G Mean	(Log_e)	ICC (2,k)	P_{Diff}
Eye Perceptual Latency	0.343	(-1.069)	0.302	(-1.198)	0.325	(-1.125)	.754	.302
Neck Perceptual Latency	0.196	(-1.632)	0.183	(-1.698)	0.178	(-1.732)	.836	.560
Arm Perceptual Latency	0.204	(-1.587)	0.152	(-1.887)	0.153	(-1.877)	.763	.101
Step Perceptual Latency	0.209	(-1.565)	0.173	(-1.752)	0.175	(-1.746)	.724	.200
Eye Response Time	0.569	(-0.565)	0.636	(-0.453)	0.626	(-0.469)	.468	.563
Neck Response Time	0.228	(-1.477)	0.183	(-1.696)	0.185	(-1.686)	.796	.033
Arm Response Time	0.198	(-1.617)	0.166	(-1.797)	0.155	(-1.862)	.701	.023
Step Response Time	0.202	(-1.601)	0.170	(-1.770)	0.173	(-1.756)	.693	.233

G Mean: Geometric Mean (Estimated Intra-Individual Variability Median of Original Data [Back-Transformation of Log_e Value])
ICC (2,k): Two-Way Random Effects, Absolute Agreement, Average of Measures Intraclass Correlation Coefficient
P_{Diff}: P-value for Repeated Measures Analysis of Variance Difference among Sessions

17

Table 6. Pairwise Comparisons of Natural Log Transformation of Original Intra-Individual Variability Value: Test-Retest at 24-hour Intervals; n=19 Graduate Students

	Session 1 – Session 2					Session 2 – Session 3				
Metric	ICC (2,k)	P_{Diff}	%Diff	(95% CI)	%CV	ICC (2,k)	P_{Diff}	%Diff	(95% CI)	%CV
Eye Perceptual Latency	.883	.006	13.8	(4.4, 24.0)	19.6	.538	.462	-7.0	(-24,2, 14.0)	52.7
Neck Perceptual Latency	.881	.353	6.8	(-7.6, 23.5)	35.3	.744	.749	3.5	(-17.1, 29.1)	58.3
Arm Perceptual Latency	.736	.017	34.9	(6.3, 13.6)	63.9	.733	.935	-1.0	(-22.5, 26.6)	66.4
Step Perceptual Latency	.744	.046	20.7	(0.4, 45.1)	46.6	.747	.949	-0.7	(-20.4, 23.9)	58.2
Eye Response Time	.347	.399	-10.6	(-31.9, 17.4)	75.8	.389	.861	1.6	(-16.2, 23.3)	49.3
Neck Response Time	.792	.009	24.5	(6.4, 45.6)	38.5	.706	.919	-1.0	(-19.5, 21.7)	53.6
Arm Response Time	.781	.020	19.7	(3.2, 38.8)	35.9	.493	.501	6.7	(-12.5, 30.0)	50.8
Step Response Time	.819	.015	18.3	(3.7, 35.0)	31.4	.579	.909	-1.3	(-22.6, 25.7)	65.3

ICC (2,k): Two-Way Random Effects, Absolute Agreement, Average of Measures Intraclass Correlation Coefficient
P_{Diff}: P-value for Paired Samples t-Test Difference between Sessions
%Diff: Symmetric Percentage Difference between Sessions (95% Confidence Interval)
%CV: Percentage Coefficient of Variation between Sessions

18

Natural Log Transformation

- 40-Trial Mean Values for 8 VR Metrics
  - ICC Test-Retest Reliability
    - Sessions 1-2-3: 4 Excellent; 3 Good; 1 Moderate
    - Sessions 1-2: 2 Excellent, 4 Good, 1 Moderate, 1 Poor
    - Sessions 2-3: 4 Excellent, 3 Good, 1 Moderate
- 40-Trial Intra-Individual Variability Values for 8 VR Metrics
  - ICC Test-Retest Reliability
    - Sessions 1-2-3: 2 Good, 5 Moderate, 1 Poor
    - Sessions 1-2: 4 Good, 3 Moderate, 1 Poor
    - Sessions 2-3: 6 Moderate, 2 Poor

Table 7. Rate Correct Score (RCS) Mean Values Test-Retest at 24-hour Intervals; n=19 Graduate Students

	Session 1	Session 2	Session 3	Sessions 1-2-3				
Composite Metric	Mean ± Std Dev	Mean ± Std Dev	Mean ± Std Dev	ICC (2,k)	P_{Diff}	S_p	SEM	MDC_{95}
RCS – Perceptual Latency	1.28 ± 0.26	1.44 ± 0.24	1.46 ± 0.23	.887	.120	.253	.085	.289
RCS – Response Time	0.75 ± 0.13	0.82 ± 0.11	0.84 ± 0.10	.851	<.001	.118	.046	.155

ICC (2,k): Two-Way Random Effects, Absolute Agreement, Average of Measures Intraclass Correlation Coefficient
P_{diff}: P-value for Repeated Measures Analysis of Variance Difference among Sessions
S_p: Pooled Standard Deviation
SEM: Standard Error of Measurement
MDC_{95}: Minimum Detectable Change at 95% Level of Confidence

Table 8. Pairwise Session Comparisons of Rate Correct Score (RCS) Mean Values Test-Retest at 24-hour Intervals; n=19 Graduate Students

	Sessions 1-2					Sessions 2-3				
Composite Metric	ICC (2,k)	P_{Diff}	s_p	SEM	MDC_{95}	ICC (2,k)	p_{Diff}	s_p	SEM	MDC_{95}
RCS – Perceptual Latency	.837	<.001	.259	.105	.290	.925	.391	.232	.064	.177
RCS – Response Time	.805	.002	.122	.054	.149	.900	.201	.105	.033	.092

ICC (2,k): Intraclass Correlation Coefficient, Two-Way Random Effects, Absolute Agreement, Average of Measures
p_{Diff}: P-value for Repeated Measures Analysis of Variance Follow-up Test for Difference between Sessions
s_p: Pooled Standard Deviation
SEM: Standard Error of Measurement
MDC_{95}: Minimal Detectable Change at 95% Level of Confidence

21

Discussion

- Session-to-session reliability of most VR mean values good to excellent
  - Substantial change (improvement) evident from session 1 to session 2
  - Lesser change observed from session 2 to session 3
- Session-to-session reliability of most VR IIV values moderate to good
  - Substantial change (improvement) evident from session 1 to session 2
  - Lesser change observed from sessions 1-2 than from sessions 2-3
- Session 1-2 Rate Correct Score (PL and RT) reliability excellent
  - Distribution normality confirmed (no transformation)
  - Composite metrics reflect both speed and accuracy

22

Clinical Relevance

- VR test assessed integrated eye, neck, arm, and whole-body responses rather than a single isolated response[3]
- RCS (correct responses per second of cumulative trial completion time) has excellent reliability as an indicator of speed/accuracy trade-off
  - RCS values did not require Log_e transformation to improve normality
  - Minimal detectable change (95% confidence) between Sessions 2-3:
    - Perceptual Latency RCS MDC_{95} = .177
    - Response Time RCS MDC_{95} = .092
- Perceptual Latency RCS (primarily visual-cognitive processes) demonstrated slightly better reliability than Response Time RCS

Clinical Relevance

- To ensure good to excellent reliability of VR 40-trial mean values (Sessions 2-3), a thorough test orientation should be provided
- Adequate reliability of VR 40-trial IIV values evident between Sessions 1-2
  - Despite inability to calculate MDC_{95} for Log_e transformed data, low IIV values relate to efficiency of perceptual-motor processes in the brain
- Immersive VR appears to provide reliable measurements of perceptual-motor function that have the potential to identify impaired brain processes, which might otherwise remain undetected

References

1. Lynall RC, Johnson RS, Lempke LB, Schmidt JD. Test–retest reliability of a functional reaction time assessment battery [published online ahead of print May 5, 2021]. *J Sport Rehabil*. 2021;doi: 10.1123/jsr.2021-0021.
2. Fjell AM, Westlye LT, Amlien IK, Walhovd KB. Reduced white matter integrity is related to cognitive instability. *J Neurosci*. 2011;31(49):18060-18072.
3. Santos, FV, et al. , Virtual reality in concussion management: from lab to clinic. *J Clin Transl Res*, 2020;5(4):148-154.
4. Grady CL, Garrett DD. Brain signal variability is modulated as a function of internal and external demand in younger and older adults. *Neuroimage*. 2018;169:510-523.
5. Wilkerson GB, Colston MA, Acocello SN, Hogg JA, Carlson LM. Subtle impairments of perceptual-motor function and well-being are detectable among military cadets and college athletes with self-reported history of concussion. *Front Sports Act Living*. 2023;5:1046572. doi:10.3389/fspor.2023.1046572
6. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163.