

8.2 Test-Retest Reliability

The present study was planned to generate repeated-measures data (i.e., measurements were taken on each participant under each experimental condition for several times). The study design makes it possible to consider the question of reliability, that is, how stable the position of a given score in a distribution of scores was when measured at different times. First of all, the stability coefficients were computed for each component task performed separately under two difficulty conditions (see Table B1).

Table B1. *Stability of Measurement Between Single-Task Sessions as a Function of Task, Sample, and Difficulty Condition*

Task	Sample	Easy				Difficult			
		1/2	2/5	5/7	1/7	3/4	4/6	6/8	3/8
RT	Total	.77**	.52**	.71**	.65**	.91**	.94**	.88**	.87**
	Young	.80**	.67**	.75**	.63**	.82**	.75**	.36	.55*
	Old	.66**	.20	.32	.36	.82**	.88**	.80**	.77**
Balance	Total	.76**	.71**	.83**	.78**	.85**	.87**	.87**	.86**
	Young	.82**	.68**	.64**	.55*	.65**	.61**	.66**	.65**
	Old	.41	.33	.67**	.58*	.75**	.81**	.75**	.78**

Note. * $p < .05$, ** $p < .01$.

In the total sample, stability coefficients between the single-task measurements of the balance performance under the easy condition (Sessions 1, 2, 5, and 7) ranged from $r = .71^*$ (Sessions 2/5) to $r = .83^{**}$ (Sessions 5/7). In the difficult condition (Sessions 3, 4, 6, and 8), the range was from $r = .85^*$ (Sessions 3/4) to $r = .87^{**}$ (Sessions 6/8). A similar pattern was observed for the reaction-time task. In the easy condition, the coefficients were lower (from $r = .52^*$ (Sessions 2/5) to $r = .77^{**}$ (Sessions 1/2) than in the difficult condition (from $r = .87^*$ (Sessions 6/8) to $r = .94^{**}$ (Sessions 4/6). Note that the correlations between Sessions 5/7 and 6/8 are high. The single-task performance assessed

in these sessions was used for the computation of the baseline level for the dual-task performance. As the high intercorrelations could be due to the extremely heterogeneous sample, all stability coefficients are displayed separately for young and old adults (see Table B1). It can be seen that in some sessions the subsamples differ in the degree of performance stability. The main conclusion drawn from the data on test-retest reliability for the single-task condition is that, in both domains, less stability was found in the easy than in the difficult condition (see Table B1).

I also examined the reliability of the dual-task performance. In the dual-task blocks of this study (Session 2 and 5; Session 4 and 6), research participants were asked to perform the component tasks concurrently under three different instructions (see Section 4.2.2.3). Thus, it was necessary to check the test-retest reliability for the performance under the task-priority instructions. As can be seen in Table B2, the reliability ranged from acceptable to high under all instructions. However, similar to the findings on stability for the single-task performance, age groups differed in several correlations.

Table B2. *Stability of Measurement Between Dual-Task Sessions as a Function of Task, Sample, Instruction, and Difficulty Condition (Raw Scores)*

Task	Sample	Focus on Balance		Focus on RT		Equal Emphasis	
		Easy	Difficult	Easy	Difficult	Easy	Difficult
RT	Total	.65**	.94**	.88**	.96**	.78**	.93**
	<i>Young</i>	.38	.82**	.75**	.77**	.64**	.68**
	<i>Old</i>	.58*	.83**	.72**	.93**	.60**	.84**
Balance	Total	.85**	.89**	.86**	.84**	.88**	.89**
	<i>Young</i>	.66**	.85**	.77**	.32	.44	.72**
	<i>Old</i>	.70**	.62**	.65**	.55*	.82**	.72**

Note. * $p < .05$, ** $p < .01$.

As analyses of dual-task data were run on two levels, the level of the raw scores and the DTCs, the stability coefficients for proportional costs were also examined. Table B3 shows that the intercorrelations of DTCs were generally considerable, except for some instructional conditions (e.g., stability of DTCs in the balance task under the instruction “Focus on Balance”). In general, the level of reliability of the DTCs was lower than that of raw scores, most probably because the dual-task costs are difference scores. Several authors pointed out that difference scores are known for their insufficient reliability (see Cohen & Cohen, 1983; Maden, Pierce, & Allen, 1993; Wittmann, 1988).

Table B3. *Stability of Measurement Between Dual-Task Sessions as a Function of Task, Sample, Instruction, and Difficulty Condition (DTCs)*

Task	Sample	Focus on Balance		Focus on RT		Equal Emphasis	
		Easy	Difficult	Easy	Difficult	Easy	Difficult
RT	Total	.48**	.68**	.33*	.67**	.25	.47**
	<i>Young</i>	.37	.82**	.48*	.60**	.36	.08
	<i>Old</i>	.60*	.54**	.25	.66**	.22	.48*
Balance	Total	.57**	.25	.57**	.35*	.55**	.25
	<i>Young</i>	.20	.52*	.77**	.69**	.18	.44
	<i>Old</i>	.65**	-.04	.65**	.48*	.74**	.05

Note. * $p < .05$, ** $p < .01$.