

# Wonderlic Basic Skills Test



# Technical Manual

Wonderlic, Inc. 400 Lakeview Parkway, Suite 200 Vernon Hills, IL 60061 800.323.3742 • Fax 847.680.9492 www.wonderlic.com

TABLE OF CONTENTS         Introduction	4	Ł
--	---	---

Technical Information	5
ATB Minimum Score Determination	5
Extended Time Test Score Adjustment 1	1
Test Validity	3
Test Content and Student Performance	4
Grade Level Equivalency	6
Test Reliability and Errors of Measurement 20	0
Item Response Theory 2!	5
Test Development	1
Acknowledgements 42	2

References	 43

Wonderlic, Inc. 400 Lakeview Parkway Suite 200 Vernon Hills, IL 60061 800.323.3742 Fax 847.680.9492 www.wonderlic.com

© 2011 Wonderlic, Inc. Unauthorized reproduction of any portion of this manual is a violation of copyright laws.

Print 4/11

# INTRODUCTION

The Wonderlic Basic Skills Test (WBST) is a short form measure of adult language and math skills which are generally learned in high school. The content of the WBST measures levels of General Educational Development (GED) as defined by the U.S. Department of Labor in the *Dictionary of Occupational Titles*. As a measure of basic language and math skills, the WBST can serve many purposes. The WBST is commonly used to help identify students and job applicants who have the skills to successfully handle the written and computational requirements of any career training program or entry level job. The WBST is used by schools, government agencies, industry, and private businesses in selecting applicants who are likely to succeed. The WBST is also used to evaluate improvement in skills as a result of educational training and relevant experiences.

On October 25, 1996, the WBST was approved by the U.S. Department of Education for use in qualifying non-high school graduates to receive Federal financial assistance for postsecondary training under Title IV of the Higher Education Act of 1965. As recommended, these students must be tested to determine whether they have the ability to benefit from postsecondary school training programs. This testing has become known as "Ability-to-Benefit" or ATB testing.

The Wonderlic ATB Program was developed to serve postsecondary schools that purchase the WBST for use in ATB testing. The program institutes testing procedures that fulfill all requirements of the U.S. Department of Education for the approved use of the WBST. These procedures are described in the Wonderlic Basic Skills Test User's Manual for Ability-to-Benefit Testing (Wonderlic, 2011). The technical manual that you are now reading was designed to accompany the aforementioned manual, providing information concerning the psychometric properties of the WBST and how it was developed. This technical manual provides detailed information concerning various aspects of test performance within ATB environments.

Thank you for selecting the WBST and the Wonderlic ATB Program. Your participation in this program provides a valuable service to students. Since 1937, over 100 million individuals have taken Wonderlic assessments in seeking new careers, training, and school admissions. We look forward to working with you as a valued partner in the Wonderlic ATB Program.

If you have any questions or concerns, please do not hesitate to call Wonderlic Education Solutions at (877) 605-9494.

# TECHNICAL INFORMATION

# **ATB Minimum Score Determination**

The original ATB minimum scores for both the verbal and quantitative sections of the WBST were prescribed by the U.S. Department of Education in accordance with the following regulations:

"...to demonstrate that a test taker has the ability to benefit from the education and training offered, the Secretary specifies that the passing score on each approved test is one standard deviation below the mean for students with high school diplomas who have taken the test within three years before the date on which the test is submitted to the Secretary for approval." (U.S. Department of Education, Dec. 1, 1995, Fed. Reg., §668.147)

In addition, the normative sample should be made up of:

"...a contemporary population representative of persons who are beyond the usual age of compulsory school attendance in the United States." (U.S. Department of Education, Dec. 1, 1995, Fed. Reg., §668.144)

In accordance with these regulations, Wonderlic selected an ATB normative sample which included students who had a high school diploma, who had completed exactly 12 years of education, and who were 16 to 29 years of age for the Secretary's use in prescribing the appropriate minimum scores. The appropriate age group selected for this sample was identified based on the vast majority of a contemporary student population representing 350,896 test administrations in the Wonderlic historical ATB database for the Scholastic Level Exam (SLE). The ATB normative sample included 4,790 test scores representing 3,706 students from 210 reporting organizations across 42 states. As shown in Table 1, this sample was comprised of an approximately equal proportion of males and females and represented a broad range of cultural backgrounds.

As described in the "Test Development" section of this manual, all forms of the same test section, either verbal or quantitative, have been developed on a common metric using item response theory. Therefore, scores achieved on different verbal or quantitative test forms are directly comparable. Tables 2 and 3 present score distributions for the ATB normative sample. As shown in the tables, distributions for the individual forms of the same section demonstrate no statistically significant differences in mean and standard deviation. Based on the combined score distributions in Tables 2 and 3, the Secretary prescribed a minimum verbal score of 200 and a minimum quantitative score of 210 for ATB determination. These scores represented one standard deviation below the average score - or approximately the 16th percentile - of all students in the Wonderlic ATB normative sample.

Wonderlic has revisited the WBST ATB minimum scores every 5 years since 1996. The merits of the aforementioned original minimum scores have been apparent during each review cycle, and thus these minimum scores have been retained subsequent to each review. Tables 4 and 5 present relevant normative tables from Wonderlic's 2011 review cycle for the Verbal and Quantitative tests, respectively. The column labeled Post-Secondary Program Applicants presents normative findings for US post-secondary program applicants possessing a high school diploma or GED certificate. The column labeled US Workforce (Matching BLS) presents normative findings from an employer sample containing US job applicants and current employees with high school diplomas or GED certificates. The latter normative sample weighted individual cases to achieve the demographic proportions associated with high school graduates (or equivalent) in the US civilian workforce in 2009, as indicated by the US Bureau of Labor Statistics ("BLS"). It should be noted that observed N sizes in Table 5 were weighted to reflect BLS proportions. This resulted in weighted N sizes that were in decimal form. Actual weighted N sizes differed slightly from those presented in this table, due to the rounding of these decimal values.

Wonderlic's 2011 review cycle took place within the context of extensive changes to Federal regulations involving ATB programs. Wonderlic served as an active participant in the Negotiated Rulemaking process that informed the Department of Education's October 29, 2010 Final Rule on Program Integrity Issues (34 CFR Parts 600, 602, 603, et al.). With respect to the establishment of ATB passing scores, the updated regulations in the Final Rule state the following:

Except as provided in §§ 668.144(d), 668.148, and 668.149, to demonstrate that a test taker has the ability to benefit from the education and training offered by the institution, the Secretary specifies that the passing score on each approved test is one standard deviation below the mean score of a sample of individuals who have taken the test within the three years before the test is submitted to the Secretary for approval. The sample must be representative of the population of high school graduates in the United States. (Authority: 20 U.S.C. 1091(d)) (34 CFR §668.147)

The norms presented in these two tables are consistent with the requirements of 34 CFR §668.147, focusing solely on individuals with a high school diploma or equivalent (and no further education beyond that).

		GEN	NDER		ALL			
	MA	LE	FEM/	FEMALE		UNKNOWN		
RACE	N	%	N	%	N	%	N	%
White	1,128	58.1	897	51.8	8	24.2	2,033	66.2
Black/African American	345	17.8	393	22.7	3	9.1	741	14.4
Hispanic	239	12.3	252	14.5	2	6.1	493	9.3
Latino/Mexican	99	5.1	89	5.1	1	3.0	189	2.2
Native American	5	0.3	26	1.5	0	0.0	31	0.7
Indo-Chinese	2	0.1	0	0.0	0	0.0	2	0.1
Japanese/Korean	5	0.3	3	0.2	0	0.0	8	0.1
Southeast Asian	23	1.2	8	0.5	2	6.1	33	0.4
Pacific Islander	5	0.3	0	0.0	0	0.0	5	0.1
Other	33	1.7	22	1.3	0	0.0	55	1.4
Unknown	56	2.9	43	2.5	17	51.5	116	5.1
All	1,940	52.4	1,733	46.8	33	0.9	3,706	100.0

# Table 1: Demographic Frequencies of WBST ATB Normative Group - High School Graduates, 16-29 Years of Age

Table 2: Distribution of WBST Verbal Scores, ATB Normative Group -High School Graduates, 16-29 Years of Age

	Test		FORM VS	5-1		FORM VS-2	2		COMBINE	0
	Score	N	%	Cum.%	Ν	%	Cum.%	Ν	%	Cum.%
		25	1 3	1 3	22	15	15	47	1 4	14
	10	- 23	0.2	1.4	1	0.1	1.5	4	0.1	1.5
	20	2	0.1	1.5	2	0.1	1.7	4	0.1	1.6
	30	- 3	0.2	1.7	4	0.3	2.0	7	0.2	1.8
	40	2	0.1	1.8	2	0.1	2.1	4	0.1	1.9
	50	5	0.3	2.0	5	0.3	2.4	10	0.3	2.2
	60	8	0.4	2.4	5	0.3	2.8	13	0.4	2.6
	70	6	0.3	2.8	2	0.1	2.9	8	0.2	2.8
	80	13	0.7	3.4	4	0.3	3.2	17	0.5	3.3
	90	12	0.6	4.0	8	0.5	3.7	20	0.6	3.9
	100	13	0.7	4.7	3	0.2	3.9	16	0.5	4.4
	110	15	0.8	5.5	9	0.6	4.5	24	0.7	5.0
	120	15	0.8	6.2	17	1.1	5.7	32	0.9	6.0
	130	13	0.7	6.9	15	1.0	6.7	28	0.8	6.8
	140	22	1.1	8.0	13	0.9	7.5	35	1.0	7.8
	150	25	1.3	9.3	15	1.0	8.5	40	1.2	9.0
	160	24	1.2	10.5	19	1.3	9.8	43	1.2	10.2
	170	32	1.6	12.1	19	1.3	11.1	51	1.5	11.7
	180	29	1.5	13.6	31	2.1	13.2	60	1.7	13.4
		29	1.5	15.1	35	2.4	15.5	64	1.9	15.3
inimum	200	48	2.4	17.5	21	1.4	17.0	69	2.0	17.3
= 200	210	28	1.4	19.0	32	2.2	19.1	60	1.7	19.0
	220	30	1.5	20.5	37	2.5	21.6	67	1.9	21.0
	230	41	2.1	22.6	30	2.0	23.6	71	2.1	23.0
	240	45	2.3	24.9	39	2.6	26.2	84	2.4	25.5
		45	2.3	27.2	3/	2.5	28.7	82	2.4	27.9
	260	55	2.8	30.0	36	2.4	31.2	91	2.6	30.5
	2/0	56	2.9	32.8	45	3.0	34.2	101	2.9	33.4
	280	60 60	3.1	35.9	52	3.5	37.7	112	3.2	30.7
	290	00	3.4 4 4	39.3	59	4.0	41.7	125	3.0	40.3
	210	67	4.4	43.7	57	4.2	43.0	140	4.5	44.0
	320	90	<u> </u>	51 7	ر الا	3.0	52 g	124	3.0	40.2 52.2
	330	75	3.8	55.5	47	3.2	56.0	122	3.5	55.7
	340	78	4.0	59.5	47	3.2	59.2	125	3.6	59.7
	350	84	4.3	63.7	50	3.4	62.5	134	3.9	63.2
	360	61	3.1	66.9	47	3.2	65.7	108	3.1	66.3
	370	69	3.5	70.4	52	3.5	69.2	121	3.5	69.9
	380	53	2.7	73.1	42	2.8	72.0	95	2.8	72.6
	390	59	3.0	76.1	47	3.2	75.2	106	3.1	75.7
	400	52	2.7	78.7	41	2.8	77.9	93	2.7	78.4
	410	49	2.5	81.2	37	2.5	80.4	86	2.5	80.9
	420	68	3.5	84.7	36	2.4	82.8	104	3.0	83.9
	430	48	2.4	87.1	32	2.2	85.0	80	2.3	86.2
	440	47	2.4	89.5	32	2.2	87.1	79	2.3	88.5
	450	35	1.8	91.3	36	2.4	89.6	71	2.1	90.6
	460	28	1.4	92.8	18	1.2	90.8	46	1.3	91.9
	470	24	1.2	94.0	16	1.1	91.9	40	1.2	93.1
	480	8	0.4	94.4	21	1.4	93.3	29	0.8	93.9
	<u>    490    </u>	22	1.1	95.5	12	0.8	94.1	34	1.0	94.9
	500	88	4.5	100.0	88	5.9	100.0	176	5.1	100.0
-										
Medians	in blac	k bars	Total	1961		Total	1486		Tota	3447
			Average	310**	¢.	Average	310*	*	Average	e 310*
Quartiles	in whit	e bars	St. Dev.	110*		St. Dev.	113*		St. Dev.	111*
	L		Median	320		Median	315		Median	320
		Interquar	tile Range	250-390	Interqua	rtile Range	240-390	Interqua	rtile Range	240-390

\* Variance of VS-1 and VS-2 show no significant difference ( $H_0$ : variance are equal, F=1.04, DF=(1485, 1960), Prob > F=.3832) \* Means of VS-1 and VS-2 show no significant difference ( $H_0$ : means are equal, T=-.2302, DF=(1485, 1960), Prob > ITI=.8179)

Table 3: Distribution of WBST Quantitative Scores, ATB Normative Group -High School Graduates, 16-29 Years of Age

	Test		FORM QS-	1	F	ORM QS-2	2	(	OMBINE	
	Score	N	%	Cum.%	Ν	%	Cum.%	N	%	Cum.%
	0	2	0.2	0.2	1	0.2	0.2	л Х	0.2	0.2
	10	2	0.2	0.2	T	0.2	0.2	J	0.2	0.2
	20	•	•	0.2			0.2	•	•	0.2
	30	1	0.1	0.4			0.2	1	0.1	0.3
	40	1	0.1	0.5			0.2	1	0.1	0.4
	50	-	0.1	0.5			0.2	-	0.1	0.4
	60	2	0.2	0.7	1	0.2	0.4	3	0.2	0.6
	70	2	0.2	1.0			0.4	2	0.1	0.7
	80	3	0.4	1.4	1	0.2	0.6	4	0.3	1.0
	90			1.4			0.6		3	1.0
	100	3	0.4	1.7	1	0.2	0.8	4	0.3	1.3
	110	2	0.2	2.0	1	0.2	0.9	3	0.2	1.6
	120	6	0.7	2.7	1	0.2	1.1	7	0.5	2.1
	130	10	1.2	3.9	6	1.1	2.3	16	1.2	3.3
	140	8	1.0	4.9	6	1.1	3.4	14	1.0	4.3
	150	4	0.5	5.4	4	0.8	4.1	8	0.6	4.9
	170	12	1.5	6.9	/	1.3	5.5	19	1.4	6.3 7.6
	100	14	1.7	8.0 10 5	3 10	0.0	0.0	1/	1.3	7.0
	100	13	1.0	10.5	10	1.9	10.7	23	1.9	9.5
	200	18	2.1	13.8	19	2.0	14.3	37	2.8	14.0
ATB Minimum	210	17	2.1	15.9	15	2.8	17.1	32	2.4	16.4
Score = 210	220	20	2.5	18.4	19	3.6	20.7	39	2.9	19.3
	230	20	2.5	20.8	17	3.2	23.9	37	2.8	22.0
	240	22	2.7	23.6	21	3.9	27.8	43	3.2	25.2
	250	35	4.3	27.9	18	3.4	31.2	53	3.9	29.2
	260	28	3.5	31.3	37	7.0	38.2	65	4.8	34.0
	270	33	4.1	35.4	35	6.6	44.7	68	5.1	39.1
	280	48	5.9	41.3	33	6.2	50.9	81	6.0	45.1
	290	57	7.0	48.3	35 20	0.0 5.5	57.5	92	6.9	52.0
	310	60	7.4	63.1	23	4.3	67.3	83	6.2	64.8
	320	43	53	68.4	17	3.2	70.5	60	4 5	69.2
	330	47	5.8	74.2	18	3.4	73.9	65	4.8	74.1
	340	32	3.9	78.2	20	3.8	77.6	52	3.9	78.0
	350	36	4.4	82.6	13	2.4	80.1	49	3.6	81.6
	360	32	3.9	86.6	11	2.1	82.1	43	3.2	84.8
	370	20	2.5	89.0	17	3.2	85.3	37	2.8	87.6
	380	15	1.8	90.9	10	1.9	87.2	25	1.9	89.4
	390	15	1.8	92.7	11	2.1	89.3	26	1.9	91.4
	400	4	0.5	93.2	8	1.5	90.8	12	0.9	92.3
	410	13	1.0	94.8 06.1	10	1.9	92.7	23	1./	94.0
	420	10	1.2	90.1	010	1.9	94.5	20	1.5	95.5
	440	10	1.2	98.0	5	0.9	97.2	15	1 1	97.7
	450	5	0.6	98.6	5	0.9	98.1	10	0.7	98.4
	460	5	0.6	99.3	3	0.6	98.7	8	0.6	99.0
	470	1	0.1	99.4	4	0.8	99.4	5	0.4	99.4
	480	2	0.2	99.6			99.4	2	0.1	99.6
	490			99.6			99.4			99.6
	500	3	0.4	100.0	3	0.6	100.0	6	0.4	100.0
Medians	in blac	k bars	Total	811		Total	532		Total	1343
			Average	288**		Average	286*	*	Average	287**
Quartiles	in whit	e bars	St. Dev.	78*		St. Dev.	79*		St. Dev.	78*
		Intorquar	Median	250-240	Intorque	Median rtilo Panco	280	Intorquar	rieaian tilo Panco	290
		Incryual	ine Kanye	200-040	Turcerqual	ne nanye	270-340	TUCEI qual	the range	270-340

\* Variance of QS-1 and QS-2 show no significant difference (H<sub>0</sub>: variance are equal, F=1.02, DF=(531, 810), Prob > F=.7922) \* Means of QS-1 and QS-2 show no significant difference (H<sub>0</sub>: means are equal, T=.3961, DF=(531, 810), Prob > ITI=.6921)

#### Table 4: Distributions of WBST Verbal Scores, 2008-2011 Normative Groups

Group 1: HS Education, Post-Secondary Program Applicants Group 2: HS Education, Job Applicants and Current Employees Weighted According to BLS Civilian Workforce Data

Teet	1: Post-Sec	ondary Progra	am Applicants	2: US Wo	rkforce (I	Matching BLS)
Score	Ν	%	Cum. %	Weighted N	%	Cum. %
0	477	1.1	1.1	23	0.5	0.5
10	60	0.1	1.2	0	0.0	0.5
20	86	0.2	1.4	3	0.1	0.5
30	84	0.2	1.6	2	0.0	0.6
40	111	0.3	1.9	6	0.1	0.7
50	154	0.4	2.3	8	0.2	0.9
60	176	0.4	2.7	8	0.2	1.0
70	213	0.5	3.2	12	0.3	1.3
80	288	0.7	3.8	6	0.1	1.4
90	327	0.8	4.6	21	0.4	1.8
100	376	0.9	5.5	13	0.3	2.1
110	466	1.1	6.6	20	0.4	2.5
120	538	1.3	7.8	19	0.4	2.9
130	636	1.5	9.3	23	0.5	3.3
140	709	1.7	10.9	43	0.9	4.2
150	849	2.0	12.9	52	1.0	5.2
160	1013	2.4	15.3	41	0.8	6.1
170	1087	2.5	17.8	57	1.1	7.2
180	1271	3.0	20.8	61	1.2	8.4
190	1405	3.3	24.1	70	1.4	9.8
200	1523	3.5	27.6	63	1.3	11.1
210	1813	4.2	31.8	75	1.5	12.6
220	1942	4.5	36.4	100	2.0	14.7
220	2131	5.0	41.3	95	1.9	16.6
230	2291	5.3	46.7	106	2.1	18.7
250	2345	5.5	52 1	123	2.5	21.2
260	2238	5.2	57.3	139	2.8	24.0
200	2271	5.3	62.6	142	2.9	26.9
210	2080	1.0	67.5	175	3.5	30.4
280	1800	4.9	71.0	173	3.5	33.0
290	1722	4.0	71.5	210	4.0	29.1
300	1733	4.0	70.0	210	4.2	42.7
310	1004	3.0	79.5	224	4.0	42.7
320	1300	3.0	02.0	201	4.1	40.7
330	020	2.0	00.1	201	4.1	50.8
340	930	2.2	07.3	231	4.7	50.4
350	770	1.0	89.1	1/0	3.0	59.0
360	724	1.7	90.8	193	3.9	62.9
370	5/5	1.3	92.1	150	3.1	66.1
380	503	1.2	93.3	102	3.1	72.5
390	327	1.2	94.0	104	3.3	72.0
400	401	0.9	95.5	145	2.9	/ 5.4
410	360	0.8	96.3	16/	3.4	/8./
420	308	0.7	97.0	154	3.1	81.9
430	260	0.6	97.6	134	2.1	84.6
440	211	0.5	98.1	89	1.8	86.4
450	158	0.4	98.5	110	2.2	88.6
460	143	0.3	98.8	102	2.1	90.6
470	103	0.2	99.1	66	1.3	92.0
480	01	0.1	99.2	54	1.1	93.1
490	/1	0.2	99.4	59	1.2	94.3
500	267	0.6	100.0	284	5.7	100.0
Median in	black bar	Total	42,905		Total	4,956
		Average	248		Average	328
Quartiles in	white bar	St. Dev.	89		St. Dev.	101
		Median	250		Median	330
	Intergu	artile Range	105	Interquartil	e Range	135

Interquartile Range

Interquartile Range

ATB Minimum Score = 210

#### Table 5: Distributions of WBST Quantitative Scores, 2008-2011 Normative Groups

Group 1: HS Education, Post-Secondary Program Applicants

Group 2: HS Education, Job Applicants and Current Employees Weighted According to BLS Civilian Workforce Data

Test	1: Post-Sec	ondary Progra	am Applicants	2: US Wo	rkforce (I	Matching BLS)
Score	Ν	%	Cum. %	Weighted N	%	Cum. %
0	219	0.5	0.5	17	0.3	0.3
10	38	0.1	0.6	0	0.0	0.3
20	37	0.1	0.7	1	0.0	0.3
30	67	0.2	0.8	0	0.0	0.3
40	60	0.1	1.0	3	0.1	0.4
50	97	0.2	1.2	1	0.0	0.4
60	125	0.3	1.5	4	0.1	0.5
70	159	0.4	1.9	6	0.1	0.6
80	218	0.5	2.4	5	0.1	0.7
90	324	0.8	3.1	8	0.1	0.8
100	343	0.8	3.9	6	0.1	0.9
110	510	1.2	5.1	17	0.3	1.2
120	578	1.3	6.5	18	0.3	1.6
130	792	1.8	8.3	27	0.5	2.0
140	933	2.2	10.5	52	1.0	3.0
150	1101	2.6	13.0	58	1.1	4.1
160	1328	3.1	16.1	81	1.5	5.6
170	1532	3.6	19.7	90	1.7	7.2
180	1585	3.7	23.4	107	2.0	9.2
190	1701	4.0	27.4	124	2.3	11.5
200	1795	4.2	31.6	158	2.9	14.4
210	1986	4.6	36.2	145	27	17.1
220	2083	4.9	41.0	175	3.2	20.3
230	2171	5.1	46.1	212	3.9	24.3
240	2110	4 9	51.0	211	3.9	28.1
250	2134	5.0	56.0	253	4 7	32.8
200	1882	3.0 4.4	60.4	262	4.8	37.7
200	1956	4.6	64.9	237	4.0	42.1
200	2087	4.0	69.8	270	5.0	47.0
200	22007	5.1	74.0	200	5.5	52.6
290	2500	5.8	80.8	316	5.8	58.4
210	2250	5.3	86.0	350	6.6	65.0
220	1676	3.0	80.0	205	5.4	70.5
220	1203	3.0	03.5	250	4.7	75.2
330	1293	3.0	92.9	202	4.7	79.0
340	600	2.0	95.0	202	3.7	70.9
350	209	1.0	90.4	104	2.0	01.7
300	390	0.9	97.4	101	3.0	97.0
370	344 204	0.0	90.2	00	2.2	07.0
300	137	0.0	90.0	80	1.7	00.0
390	110	0.3	00.9	00	1.0	01 7
400	60	0.0	99.Z	80	1.7	91.7 93.4
410	85	0.2	99.A	75	1.0	94 R
420	55	0.2	99.7	60	1.1	94.0
430	33	0.1	90.7 90.8	42	0.8	96.7
440	19	0.0	99.8	38	0.7	97.4
460	15	0.0	99.9	22	0.4	97.8
470	13	0.0	99.9	28	0.5	98.3
480	11	0.0	99.9	26	0.5	98.8
490	7	0.0	99.9	10	0.2	98.9
500	20	0.0	100.0	57	1.0	100.0
		0.0		5,		
Median in	black bar	Total	42,903		Total	5,411
		Average	236		Average	286
Quartiles in	white bar	St. Dev.	74		St. Dev.	80
		Median	240		Median	290
	Interqu	artile Range	105	Interquartil	e Range	95

ATB Minimum Score = 200

# **Extended Time Test Score Adjustment**

In certain situations, it may be appropriate to administer the WBST with a time extension. The extended time administration is designed to provide a more accurate skills assessment for applicants who are likely to be inaccurately assessed under the standard time limit.

Wonderlic research shows that a timed period causes applicants to take the test more seriously and motivates them to perform better. Therefore, the approved extended time WBST administration procedure combines a standard 20 minute timed period followed by a time extension on the same test form. This provides separate scores on both a timed and extended time basis.

The extended time score contains a statistical score adjustment\*. This adjustment is based on a linear regression of observed timed and extended time test scores from individuals who were accurately assessed during the timed testing period. This relationship provides a sound method of estimating the appropriate test score for individuals who are not accurately assessed in the standard 20-minute timed testing period. Figures 1 and 2 demonstrate this relationship for both the verbal and quantitative sections of the WBST. As shown in these figures, the appropriate score for an extended time verbal test administration is expected to equal approximately 83% of the unadjusted extended time score. Similarly, the appropriate score for an extended time quantitative test administration is expected to equal approximately 91% of the unadjusted extended time score. These adjustments will automatically be made by Wonderlic when your test is scored.

\*Please note: A score adjustment may not be appropriate when extended testing time is provided as a reasonable accomodation for an applicant with a disability. Please contact Wonderlic Education Solutions for more information.









# **Test Validity**

The Standards for Educational and Psychological Testing (AERA, APA & NCME, 1999; AERA, APA & NCME, 1985) indicate that test validity refers to the appropriateness, meaningfulness and usefulness of the specific inferences made from test scores. Test validity is regarded as "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (AERA, APA, & NCME, 1999, p.9). Inferences and interpretations made from WBST scores are often supported on the basis of content validity evidence. The 1985 edition of the Standards provides a succinct summary of content validity:

"Content-related evidence demonstrates the degree to which the sample of items, tasks, or questions on a test are representative of some defined universe or domain of content. The methods often rely on expert judgments to assess the relationship between parts of the test and the defined universe..."

The content domain measured by the WBST includes the fundamental verbal and quantitative skills which are required for successful performance in occupational training programs and in the workplace. All test content has been developed by subject-matter experts in accordance with the job-related language and math skills published by the U.S. Department of Labor, Employment and Training Administration in the *Dictionary of Occupational Titles* (U.S. Department of Labor, 1991). The test is also designed to take into account research conducted by the National Assessment of Educational Progress (NAEP), the Secretary's Commission on Achieving Necessary Skills (SCANS), and the National Educational Goals Panel.

In addition, normative test data from students in school demonstrate that proficiency with the academic subject matter measured by the WBST is strongly related to levels of academic achievement ranging from the 6th grade through four years of college or postsecondary education. After a comprehensive review of all test content, the U.S. Department of Education has approved the WBST as a valid measure of the necessary high school level language and math skills for successful performance in postsecondary occupational training programs.

All WBST items have been subjected to independent expert review for cultural sensitivity, as well as statistical analysis of Differential Item Functioning (DIF), to ensure that the test is culturally and gender equivalent and free of content bias.

Based on this evidence, the WBST is a content valid measure of basic skills, for use in applicant selection for any job or occupational training program which has been determined to require basic language and math skills. The content validity of the WBST is supported by a well-focused definition of purpose, expert development of test content, and independent expert review.

# **Test Content and Student Performance**

The U.S. Department of Education regularly evaluates the learned skills of US high school students and presents the findings in a report entitled The Nation's Report Card<sup>TM</sup>. In a 2009 report, the Department took a nationally representative sample of 12th grade public school students and classified these students into one of four proficiency levels based on their learned skills. The findings indicated that 64% and 75% of 12th grade students were performing at a Below Basic or Basic proficiency level for reading and math skills, respectively.

These findings are consistent with Wonderlic normative research which indicates that students are still learning and improving their proficiency with basic verbal and quantitative skills throughout high school and college. Their proficiency with the academic subject matter measured by each WBST item is indicated by the observed percentage of students who answer the test item correctly (p-value). Figures 3 and 4 illustrate these item level percentages for high school students who have taken forms VS-1 and QS-1. These figures reveal that on the majority of WBST items, high school students demonstrate item level proficiencies ranging from 24% to 89% for verbal skills and 10% to 88% for quantitative skills.

#### Figure 3





# **Grade Level Equivalency**

WBST scores can be interpreted in terms of grade level equivalency to provide a more familiar and generalized scale for measuring an applicant's overall level of basic skills. The median or average test scores of students in school are commonly used to establish *grade-equivalent* or *grade level scores* (Peterson, Kolen, & Hoover, 1989). A grade level score indicates the school year in which students typically achieve a given test score. For example, if a student achieves a grade level score of 9.5 on the WBST, he or she has demonstrated a level of basic skills proficiency which is typical of students who are tested in the middle of 9th grade.

The grade level sample includes 14,531 test scores representing 7,688 students from 170 reporting organizations across 41 states. As shown in Table 6, this sample is comprised of an approximately equal proportion of males and females and represents a broad range of cultural backgrounds. Participating students had completed varying levels of education ranging from 6th grade through 4 years of college or postsecondary school at the time they were tested. These students were tested with the WBST as part of career and educational assessments, summer job training programs, and standard employment application processes. While this normative group is not a complete national sample, it provides a sound basis for establishing grade level equivalency.

For the purposes of establishing grade level scores, the normative group for each grade includes only those students who are of the appropriate age for that grade level and who have successfully completed the previous grade. For example, the 9th grade normative group is comprised of individuals who are 14 or 15 years of age and who have reported completing eight years of education.

When test scores are normally distributed, group median and average scores will be similar. However, average test scores are more influenced by extreme scoring individuals, whereas median test scores provide a more accurate representation of the skills which have been demonstrated by the majority of a particular group. Therefore, WBST grade level scores are based on a linear regression of the median test scores of students who have completed from six to 16 years of education.

Figures 5 and 6 report verbal and quantitative WBST normative information by grade level. These figures graphically present median and average test scores for each grade. While the WBST does not measure skills generally associated with college level curriculum, these figures demonstrate the successive score gains observed at every level of education from 6th grade through four years of postsecondary school. This provides empirical evidence that students continue to improve their proficiency with the basic skills measured by the WBST throughout both high school and college.

Figure 7 reports WBST skills composite information by grade level. An applicant's skills composite is the simple arithmetic average of his or her verbal and quantitative scores. The grade level score reported on the *WBST Individual ATB Score Report* is based on the skills composite. This grade level score provides the best estimate of an applicant's overall basic skills proficiency in terms of grade level equivalency.

	GENDER					ALI	L	
	MA	LE	FEMA	<b>LE</b>	UNKNO	WN		
RACE	N	%	N	%	N	%	N	%
White	2,735	68.0	2,324	64.9	28	34.6	5,087	66.2
Black/African American	541	13.4	561	15.7	5	6.2	1,107	14.4
Hispanic	338	8.4	373	10.4	6	7.4	717	9.3
Latino/Mexican	84	2.1	85	2.4	0	0.0	169	2.2
Native American	29	0.7	22	0.6	1	1.2	52	0.7
Indo-Chinese	8	0.2	0	0.0	0	0.0	8	0.1
Japanese, Korean	5	0.1	5	0.1	0	0.0	10	0.1
Southeast Asian	13	0.3	15	0.4	2	2.5	30	0.4
Pacific Islander	6	0.2	5	0.1	0	0.0	11	0.1
Other	56	1.4	52	1.5	1	1.2	109	1.4
Unknown	210	5.2	140	3.9	38	46.9	388	5.1
All	4,025	52.4	3,582	46.6	81	1.1	7,688 1	L00.0

Table 6: Demographic Frequencies of WBST Grade Level Normative Group

Table 7 presents the WBST verbal, quantitative, and composite grade-equivalent score ranges for 6th grade through college. However, WBST test content measures basic skills generally not associated with college level curriculum. Grade-equivalent scores beyond the 12th grade indicate that an applicant's basic skills proficiency is comparable to that of students in college, but these scores do not necessarily indicate that the applicant is proficient with college level skills. For this reason, grade level scores beyond the 12th grade are not specified on the *WBST Individual ATB Score Report*.

Table 7: WBS	T Grade	Equivalent	Scores
--------------	---------	------------	--------

YEAR IN S	CHOOL	VERBAL	QUANTITATIVE	COMPOSITE
Middle	6th Grade*	193-210	214-226	207-221
School	7th Grade	211-229	227-240	222-237
	8th Grade	230-247	241-253	238-253
High	9th Grade	248-266	254-266	254-268
School	10th Grade	267-284	267-280	269-284
	11th Grade	285-302	281-293	285-300
	12th Grade	303-321	294-313	301-315
College/	1st Year	322-339	314-320	316-331
Postsecondary	2nd Year	340-358	321-333	332-347
School	3rd Year	359-376	334-346	348-362
	4th Year	377-395	347-360	363-378

\* Extrapolated from linear regression - 6th Grade test data is not yet available.





# Figure 7



## **Test Reliability and Errors of Measurement**

The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1985) specifies that test reliability refers to the degree to which test scores are free from errors of measurement. According to the *Standards*:

"A test taker may perform differently on one occasion than on another for reasons that may or may not be related to the purpose of measurement. A person may try harder, be more fatigued or anxious, have greater familiarity with the content of questions on one test form than on another, or simply guess correctly on more questions on one occasion than on another. For these and other reasons a person's score will not be perfectly consistent from one occasion to the next."

The 1999 edition of the Standards (AERA, APA, & NCME, 1999) further elucidate this notion, stating the following:

"Reliability refers to the consistency of such measurements when the testing procedure is repeated on a population of individuals or groups." (p. 25)

#### Standard Errors of Measurement

From the above statements, it is evident that a score from any particular test administration is only an estimate of an individual's actual level of knowledge, skills or ability ("true score"), and that this estimation is always subject to some error of measurement. The likely amount of error in an individual's test score is reflected in the *standard error of measurement* for the given score. An individual's true score is likely to fall within one standard error of measurement of his or her observed test score. Tests that have relatively lower standard errors of measurement have a higher degree of reliability. Table 8 provides a complete listing of standard errors for test scores on each form of the WBST.

Standard errors of measurement can be used to compare applicant test scores to the ATB minimum score or to the scores of other applicants. If an applicant's score falls below the ATB minimum score, but the minimum score falls within one standard error of his or her score, then it may be beneficial to retest the applicant. In addition, when the standard error range around one applicant's test score overlaps with that of another applicant, then their scores are not considered to be significantly different. For example, the standard error ranges for applicants scoring 170 and 190 on form VS-1 are 140-200 and 161-219 respectively. Because these ranges overlap, the applicants' scores are not considered to be significantly different.

Test		TEST	FORM	
Score	VS-1	VS-2	QS-1	QS-2
0	62	54	72	68
<u> </u>	60	52	70	65
10	58	51	67	62
15	56	49	65	59
20	54	48	63	56
25	52	46	61	54
30	50	45	59	52
35	49	44	58	50
40	48	43	56	48
45	46	42	54	46
50	45	41	53	45
55	44	40	52	43
60	43	39	50	42
65	42	38	49	41
70	41	37	48	40
75	40	37	46	39
80	39	36	45	38
85	38	35	44	37
90	37	35	43	37
95	37	34	42	36
100	36	34	41	36
105	35	33	40	35
110	35	33	39	35
115	34	33	39	34
120	34	32	38	34
125	33	32	37	34
130	33	32	36	34
135	32	31	36	33
140	32	31	35	33
145	31	31	35	33
150	31	31	34	33
155	31	30	34	32
160	30	30	34	32
165	30	30	33	32
170	30	30	33	32
175	30	29	33	31
180	29	29	33	31
185	29	29	32	31
190	29	29	32	31
195	29	28	32	31
200	28	28	32	31
205	28	28	32	31
210	28	28	32	30
215	28	28	32	30
220	28	28	32	30
225	28	28	32	30
230	28	28	32	30
235	27	28	32	30
240	27	28	31	30
245	27	28	31	30
250	CO	ntinued in	next colu	mn

Test	TEST FORM					
Score	VS-1	VS-2	QS-1	QS-2		
250	27	28	31	30		
255	27	28	30	29		
260	27	28	30	29		
265	27	29	29	29		
270	27	29	28	28		
275	28	29	27	28		
280	28	29	25	28		
285	28	30	24	27		
290	28	30	22	27		
295	28	30	21	2/		
300	29	31	20	26		
210	29	21	19	25		
215	29	32	10	23		
320	29	32	18	24		
325	29	32	18	27		
330	29	33	19	22		
335	29	33	19	22		
340	29	33	19	22		
345	29	34	20	21		
350	28	34	20	21		
355	27	34	20	21		
360	26	35	20	21		
365	26	35	20	21		
3/0	25	35	20	21		
3/5	24	30	20	20		
205	24	30	20	20		
200	24	30	20	19		
395	23	37	19	19		
400	24	37	19	19		
405	24	38	19	19		
410	25	38	19	19		
415	26	39	20	19		
420	27	39	20	19		
425	28	40	21	20		
430	29	41	22	20		
435	31	41	23	21		
440	32	42	24	22		
445	34	43	25	23		
450	38	44	20	24		
460	40	46	27	20		
465	42	40	30	29		
470	44	48	32	31		
475	47	49	34	33		
480	49	50	36	35		
485	52	51	38	37		
490	55	53	40	39		
495	57	54	43	41		
500	60	56	45	44		

Table 8: WBST Standard Errors of Measurement

#### Test-Retest Reliability

The reliability evidence provided for a particular test should support the intended use and interpretation of test scores. This evidence is generally based on two fundamental methods of estimating test reliability - test-retest correlations and measures of internal consistency.

Retesting applicants for ATB purposes is generally conducted on the same day as the initial test administration. Within a 12-month period, and in accordance with approved Wonderlic ATB procedures, retesting may be conducted with an alternate form of the WBST on the same day as the initial administration or anytime thereafter. In addition, retesting may be conducted with the same form of the WBST if at least 60 days have passed since the applicant was initially administered that form and the applicant has been administered the alternate form within the past twelve months. Figures 8 and 9 show observed increases in the verbal and quantitative test scores of applicants who were retested with a WBST form they had already taken. The curves shown in these figures represent the gain in test scores which can be expected as time increases between the first and second administrations. These curves demonstrate that the observed impact of conducting a first and second test administration using the same form diminishes as the time between administrations increases. In addition, the expected gain in test scores when the second administration takes place at least 60 days after the first is well within the standard error of measurement of the ATB minimum passing scores for both the verbal and quantitative sections. Therefore, 60 days between the first and second test administrations provides a practical standard for retesting with the same form of the WBST.

The WBST retest sample includes 3,356 test scores representing 941 students from 34 reporting organizations across 23 states. As shown in Table 9, this sample is comprised of an approximately equal proportion of males and females and represents a broad range of cultural backgrounds. Table 10 presents the appropriate test-retest reliability estimates for both alternate and same-form retesting within three relevant time intervals. In addition, Table 11 presents the test-retest reliability estimates for classifications. These estimates indicate the level of consistency with which the WBST classifies applicants as either passing - or not passing - the ATB minimum verbal and quantitative WBST scores when they are retested.

#### Internal Consistency

Measures of internal consistency such as split-half correlations are generally regarded as inflated estimates of test reliability when applied to speeded tests such as the WBST. However, a comparative analysis of the Cronbach alphas shown in Table 12 for both timed and extended time WBST administrations reveals that the 20-minute timed period does not practically or significantly impact internal consistency estimates for either the verbal or quantitative section of the test. This not only confirms the appropriateness of the 20-minute time limit, but also establishes that the timedtesting period is sufficient to measure all primary verbal and quantitative content domains assessed by the WBST with a high degree of reliability.

			GEND	DER			AL	L
	MA	LE	FEM	ALE	UNKNO	OWN		
RACE	N	%	N	%	N	%	N	%
White	260	56.6	185	45.5	5	6.7	450	47.8
Black/African American	119	25.9	173	42.5	1	1.3	293	31.1
Hispanic	31	6.8	20	4.9	1	1.3	52	5.5
Latino/Mexican	8	1.7	5	1.2	0	0.0	13	1.4
Native American	5	1.1	4	1.0	0	0.0	9	1.0
Japanese/Korean	0	0.0	2	0.5	0	0.0	2	0.2
Southeast Asian	6	1.3	1	0.3	1	1.3	8	0.9
Pacific Islander	1	0.2	0	0.0	0	0.0	1	0.1
Other	8	1.7	1	0.3	0	0.0	9	1.0
Unknown	21	4.6	16	3.9	67	89.3	104	11.1
All	459	48.8	407	43.3	75	8.0	941	100.0

Table 9: Demographic Frequencies of WBST Test-Retest Sample

Median education = 12 years, Average age = 33 years.

#### Table 10: WBST Test-Retest Reliabilities Estimates

TEST		DAY OF RETEST ADMINISTRATION					
SECTION	TYPE	Same Day Within 30 Days		After 30 Days	OVERALL		
Verbal	Same Form	.90	.84	.90	.89		
Skills		(N=50)	(N=54)	(N=129)	(N=233)		
	Parallel Form	.93	.89	.92	.92		
		(N=56)	(N=103)	(N=469)	(N=628)		
	Combined	.93	.87	.92	.91		
		(N=106)	(N=157)	(N=598)	(N=861)		
Quantitative	Same Form	.83	.85	.91	.88		
Skills		(N=45)	(N=45)	(N=133)	(N=223)		
	Parallel Form	.89	.88	.83	.84		
		(N=48)	(N=82)	(N=465)	(N=596)		
	Combined	.89	.86	.86	.86		
		(N=92)	(N=127)	(N=598)	(N=817)		

All reported reliabilities are significant at p<.0001.

TEST		DAY OF RE			
SECTION	ТҮРЕ	Same Day	Within 30 Days	After 30 Days	OVERALL
Verbal	Same Form .88		.99	.87	.89
Skills		(N=50)	(N=54)	(N=129)	(N=233)
	Parallel Form	.99	.92	.89	.91
		(N=56)	(N=103)	(N=469)	(N=628)
	Combined	.93	.99	.89	.91
		(N=106)	(N=157)	(N=598)	(N=861)
Quantitative	Same Form	.89	.90	.91	.90
Skills		(N=45)	(N=45)	(N=133)	(N=223)
	Parallel Form	.89	.78	.88	.87
		(N=48)	(N=82)	(N=465)	(N=596)
	Combined	.89	.82	.89	.88
		(N=92)	(N=127)	(N=598)	(N=817)

Table 11: WBST Test-Retest Reliability Estimates for ATB Status Classification

All reported reliabilities are significant at p<.0001 and are based on tetrachoric correlations and ATB minimum passing scores for the WBST.

#### Table 12: WBST Internal Consistency Estimates (Cronbach Alphas) by Content Domain, Test Form and Test Time

	VS	-1	VS	-2	QS	-1	QS	-2	
CONTENT DOMAIN	Timed	Extended Time	Timed	Extended Time	Timed	Extended Time	Timed	Extended Time	CONTENT DOMAIN
	(N=24,003)	(N=457)	(N=11,281)	(N=396)	(N=23,124)	(N=418)	(N=11,355)	(N=354)	
Word Knowledge	.81	.83	.80	.87	.77	.89	.77	.90	Basic Computation
Sentence Construction	.89	.90	.90	.90	.83	.86	.88	.88	Quantitative Evaluation
Using Information	.83	.86	.87	.86	.82	.82	.89	.85	Algebra & Geometry
Combined	.94	.96	.95	.96	.93	.93	.93	.93	Combined

All reported reliabilities are significant at p<.0001.

# **Item Response Theory**

The WBST has been professionally developed using item response theory (IRT). The following section provides an overview of three salient concepts in IRT: 1) unidimensionality, 2) test item characteristics, and 3) test information. A basic knowledge of these concepts will help in understanding the advantages that IRT has provided in the construction of the WBST and the evaluation and interpretation of WBST scores.

#### Unidimensionality

IRT models include a set of assumptions about the data to which the model is applied. A common assumption of IRT models is that a single overall trait is being measured by the items that make up the test. When this is true, the test is said to be unidimensional. The assumption of unidimensionality cannot be strictly met because several cognitive, personality, and test-taking factors always affect test performance (Hambleton, Swaminathan & Rogers, 1991). These factors can include motivation, test anxiety or cognitive skills beyond what is measured in the test. However, the unidimensionality assumption does require the presence of a dominant factor that determines test performance. Of course, the dominant factor being measured by the WBST is basic verbal or quantitative skills.

A comparison of eigenvalues can be studied to determine whether a dominant factor is being measured by a particular test (Reckase, 1979). An eigenvalue is equal to the amount of variance in test scores that can be explained by the factor being measured. The largest eigenvalue should represent a much larger percentage of the total variance than does the next largest eigenvalue. Figure 10 shows the eigenvalue plots for WBST forms VS-1, VS-2, QS-1, and QS-2.



Figure 10 Eigenvalue Graphs for the WBST\*

\* All eigenvalues have been computed using tetrachoric correlations.

It is clear from the eigenvalue plots in Figure 10 that each form of the WBST measures one dominant trait, as is required by the unidimensionality assumption of IRT.

#### **Test Items Characteristics**

All measurements made on the basis of IRT build upon a fundamental relationship between test item performance and the specific trait being measured, such as basic quantitative or verbal skills. IRT specifies that this relationship can be described for individual test items by a mathematical function referred to as an item characteristic curve (ICC). An ICC specifies that as the level of a given trait increases, the likelihood of a correct response to a test item designed to measure that trait increases. For example, this relationship is evident in the percent-correct frequency graphs presented in Figure 11 for WBST verbal items 26 and 36.





In Figure 11 above, we see that as verbal scores increase, the percent frequency of correct responses to items 26 and 36 on WBST form VS-1 also increases.

While many mathematical forms exist for item response models, the WBST uses a three parameter logistic model. The actual three parameter logistic ICCs for items 26 and 36 on WBST form VS-1 are shown in Figure 12 below.

Figure 12





The model stipulates that the following three test item characteristics uniquely define the ICC of each item and therefore describe the way in which the item functions:

#### 1. Difficulty (Threshold)

The difficulty characteristic of an item is defined as the skills level which is necessary to demonstrate approximately a 50% likelihood of achieving a correct response to the item (point b in Figure 12). The difficulty level of a test item corresponds to the point of inflection on the ICC. The point of inflection is the steepest point on the curve and is generally at a skills level slightly higher than the point of 50% likelihood.

#### 2. Discrimination (Slope)

The discrimination characteristic of a test item is defined as its reliability in identifying and separating examinees into their appropriate skills levels. The point of maximum discrimination on an ICC (point a in Figure 12) is at the point of inflection on the curve. Items with steep slopes at the inflection point of the ICC are more useful than those with flat slopes in reliably discriminating between examinees.

#### 3. Guessing (Asymptote)

The guessing characteristic of a test item is defined as the likelihood that examinees with the lowest skills levels will achieve a correct response to the item (point c in Figure 12). In other words, by simply guessing, all examinees have at least some probability of correctly responding to the item. The guessing value of multiple choice test items, such as those on the WBST, is approximately equal to 1 divided by the number of possible choices. It follows that open-ended test items have a guessing value approximately equal to zero.

In Figure 12 we see that between the two WBST items a higher verbal skills level is necessary on item 36 to demonstrate a 50% likelihood of achieving a correct response. This means item 36 is more difficult than item 26. Also, the ICCs have similar slopes at their steepest points. This means the items have comparable capacity to separate examinees into their appropriate skills levels. Finally, people with very modest verbal skills have a slightly higher likelihood of achieving a correct response to item 26 than to item 36. This means item 26 has a higher guessing value than item 36.

The method that has been used to estimate the difficulty, discrimination, and guessing values of each item on the WBST is an iterative procedure referred to as marginal maximum likelihood estimation (MML). The steps involved in MML are beyond the scope of this discussion. For a detailed discussion of item parameter estimation procedures, refer to Hambleton and Swaminathan (1985). Nevertheless, it is quite clear that the actual percent-correct frequency graphs in Figure 11 are consistent with the theoretical ICCs which have been estimated for the WBST items.

#### **Item Information**

The ICC provides a powerful method for describing, selecting, and comparing the reliability of test items. This method involves the use of an item information function. The information function of a test item is derived from the ICC and is related to the amount of discrimination that the item provides among examinees across the full range of the trait being measured. Figure 13 shows the item information function and the ICC for WBST item 26 on form VS-1. For each test item, the maximum amount of information is provided at a level of skill which is slightly above the difficulty level of the item. In Figure 13, this skill level corresponds to point d on the ICC.

Figure 13 Item Characteristic Curves and Information Function -Item 26 Form VS-1



In Figure 13 we see that item 26 provides maximum information at the skills levels corresponding to a score of approximately 210.

#### **Test Information**

The information provided by a test is simply the sum of all item information. The test information function is calculated by vertically adding the item information functions at each point across the full range of the skills being measured. Figure 14 shows the information functions for the overall test as well as each of the GED subscales on form VS-2.

In Figure 14 we see that total test information is the vertical sum of the subscale information. It is also clear in Figure 14 that the WBST GED levels provide information at three different ranges of verbal skills. For further discussion on the GED scales, see the "Skills Standards" subsection in the "Test Development" section of this manual.



Figure 14 Test Information Functions by GED Level and Overall - Form VS-2

IRT has provided several advantages in WBST construction and in the evaluation and interpretation of WBST scores.

#### **Test Construction**

As discussed in the previous section, a key element in IRT is the item characteristic curve, which quantifies important test item characteristics such as difficulty, discrimination, and guessing. An important advantage of IRT in test construction is that these item characteristics can be applied to groups outside the group that was used to determine them. This means sample dependent group differences have no impact on the construction of test forms. This is not the case in the construction of traditional test forms, which are said to be group-dependent. The WBST can be applied equivalently to all subgroups.

Another important advantage of IRT in test construction is that the contribution that individual test items can make to the overall discrimination of the test can be identified before a test is assembled. This is because the information provided by an item is determined independently of other items, and the information provided by a test is simply the sum of all item information. Therefore, sets of items that provide discrimination across predetermined skills levels can be selected in constructing a test. This is not the case with traditional tests, which do not define item discrimination or reliability independent of other items, but rather, in reference to other items on the test.

As a test of basic skills, the WBST is not designed to provide discrimination among examinees with extreme skills levels. Some examinees have extremely modest skills and are not suited for any type of independent work. Others may have extremely strong skills and are suited for very complex jobs. These types of examinees have skills levels which are outside the range of skills accurately measured by the WBST. Therefore, all examinees who achieve scores outside this range are assigned either a score of 0 or 500.

#### **Test Scoring**

The most fundamental benefit of test scores which are based on IRT is that score calculations are made at the individual item level. This has several important advantages in test scoring.

First, IRT scores take into account more information than do scores which are based only on the number correct. Difficulty, discrimination, and guessing characteristics of not only those items answered correctly, but also those not answered correctly, are direct components of IRT score calculations. This is not the case with classical test scores which are often based only on the number correct. This additional information improves the reliability and validity of WBST scores.

Next, IRT scores and their interpretation do not depend on the particular set of test items which are administered. This means test form differences have no impact on the interpretation of test scores. Test score comparisons can be made not only for examinees who took the same (or parallel) test forms, but also for those who took different (or unparallel) test forms. This is not the case with classical test scores which are said to be test-dependent. WBST scores can be directly compared across all current and future verbal or quantitative forms. Also, the classical concept of test-retest reliability relies on the existence of parallel or equivalent test forms. However, classically equivalent measures are difficult to achieve in practice because examinees are never in the exact same state of mind during two different test administrations. New skills can be developed, motivation and anxiety levels can change, or information can simply be forgotten between administrations. Because IRT scores are not test-dependent, test-retest reliability can be readily determined across different forms of the WBST.

Finally, IRT scores have more accurate standard errors of measurement than do classical scores. Standard errors of measurement are used to make comparisons of examinee test scores. A discussion of standard errors is provided in the "Test Reliability and Errors of Measurement" section of this manual. IRT provides standard errors that are specific to each score, while traditional test scores assume that the standard error is the same for all examinees. Therefore, comparisons between WBST scores can be made with improved accuracy.

# **Test Development**

#### **Definition of Purpose**

Since 1937, over 100 million applicants have taken Wonderlic assessments in seeking new careers, training, and school admissions. During the late 1980's, Wonderlic clients began to express a growing need to identify students and job applicants who have the necessary high school level language and math skills to successfully handle the fundamental written and computational requirements of career training programs and entry level jobs. As noted in the *Dictionary of Occupational Titles*, 1991 Revised Edition:

"In the 14 years since the release of the Fourth Edition of the Dictionary of Occupational Titles (DOT), the American workplace has undergone a revolutionary change. The skills most in demand are not what they were 14 years ago; educational requirements have steadily increased. Too many of America's young people are entering the world of work inadequately prepared. The resulting dislocation - the so called 'skills gap' - presents those of us who prepare, hire or support American workers with a serious challenge."

> Robert T. Jones Assistant Secretary for Employment and Training U.S. Department of Labor

Wonderlic conducted a comprehensive review of existing language and math skills tests. It was determined that the vast majority of these assessments were too lengthy to be practically administered in business and training school environments or too focused on specific types of skills to be broadly appropriate. Also, many tests measured explicit skills knowledge without assessing the applied use of those skills. As a result, in 1991 Wonderlic began a three year development process to create a test of verbal and quantitative skills.

Working directly with business and postsecondary schools, Wonderlic determined that the most appropriate test of basic language and math skills for use in occupational assessment would: 1) contain academic subject matter that is based directly on job requirements, 2) require less than one hour for administration, and 3) be easily administered, scored, and interpreted by competent office staff.

In July of 1992, working with Dr. Nambury Raju, Professor of Psychology at the Illinois Institute of Technology, Wonderlic developed an ambitious and well-focused strategy to complete WBST development within two years time.

#### Validity of Test Content

The Standards for Educational and Psychological Testing (AERA, APA & NCME, 1999; AERA, APA & NCME, 1985) indicate that test validity refers to the appropriateness, meaningfulness and usefulness of the specific inferences made from test scores. Test validity is regarded as "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (AERA, APA, & NCME, 1999, p9). Inferences and interpretations made from WBST scores are often supported on the basis of content validity evidence. The 1985 edition of the Standards provides a succinct summary of content validity:

"Content-related evidence demonstrates the degree to which the sample of items, tasks, or questions on a test are representative of some defined universe or domain of content. The methods often rely on expert judgments to assess the relationship between parts of the test and the defined universe..."

The content domain that the WBST was designed to measure includes the fundamental verbal and quantitative skills which are required for successful performance in occupational training programs and in the workplace.

#### Skills Standards

Although test content of the WBST would be based primarily on job requirements, standards for grade level equivalency were also reviewed to provide a more familiar and generalized scale for measuring an applicant's overall level of basic skills. Wonderlic conducted an investigation of state requirements for language and math skills by grade level. However, because academic subject matter generally spans more than a single grade level, these requirements provided only illustrative information which broadly described educational standards. Skills are often introduced at one grade level and then mastered over the next two or three years. As a result, practical standards for grade level equivalency must be determined normatively based on the average or median skill level of students in a particular grade.

The skills standards for test content of the WBST would not include specific job knowledge but rather the fundamental verbal and quantitative skills upon which specific job knowledge would be built. Therefore, the WBST would be designed to measure the job-related language and math skills defined for the General Educational Development (GED) scales published by the U.S. Department of Labor in the *Dictionary of Occupational Titles* (DOT). The GED scale measures six levels of skills complexity which describe the language and math requirements of approximately 13,000 job titles. These skills have been culled from approximately 75,000 on-site job analyses conducted by the regional Occupational Analysis Field Centers of the U.S. Department of Labor.

In 1991, the Wonderlic Scholastic Level Exam (SLE) was approved by the U.S. Department of Education for use in ATB determination. ATB minimum passing passing scores for the SLE were determined for each occupational education program based directly on the GED scale. Earlier research involving the General Aptitude Test Battery (GATB) published by the U.S. Employment Service identified a strong statistical relationship between ATB minimum scores for the SLE and occupational minimum test scores for the GATB (Wonderlic, Long, & Rafilson, 1990). This research provided confirming evidence supporting the use of the GED scale as a measure of job-related language and math skills.

Although GED levels 1-6 describe the complete range of job-related skills complexity, levels 4-6 involve academic subject matter generally associated with college level curriculum, while levels 1-3 describe the basic skills required for 89.7% of the occupational titles published in the DOT. Therefore, the WBST would be designed to measure the first three levels of the GED scale.

#### Test of Verbal Skills

Questions on the WBST Test of Verbal Skills are presented in three different formats – explicit, applied, and interpretive. Explicit questions require knowledge of word meanings, grammar, and sentence construction. Applied questions involve general reading comprehension and knowledge of word meanings, grammar, and sentence construction in the context of passages involving practical applications. Interpretive questions require the understanding and use of information in diagrams, charts, tables, and graphs. The WBST Verbal Form contains 50 questions which help measure a broad range of language skills across the three primary content domains shown in the table below.

Verbal Skills Domain	Skills
Locate, Understand, and Use Information in Multiple Formats	<ul> <li>Locate specific information in various formats</li> <li>Understand and use "how to" instructions</li> <li>Use literature and other written information</li> </ul>
Recognize Word Meanings by Definition or Context	<ul> <li>Complete a sentence with an appropriate word in context</li> <li>Recognize word meanings</li> <li>Recognize multiple meanings of words used in context</li> <li>Recognize meanings of unfamiliar words used in context</li> </ul>
Recognize and Identify Proper Grammar and Sentence Construction	<ul> <li>Identify a complete sentence</li> <li>Identify agreement of subject and verb</li> <li>Recognize standard sentence structure</li> <li>Complete compound and complex sentences</li> <li>Recognize grammatical errors in compound and complex sentences</li> </ul>

#### Test of Quantitative Skills

Questions on the WBST Test of Quantitative Skills are presented in three different formats - explicit, applied, and interpretive. Explicit problems require direct math computation, quantitative evaluation, and algebra. Applied problems involve identifying and performing appropriate math operations in the context of practical applications. Interpretive problems require the evaluation and use of quantities presented in diagrams, charts, tables, and graphs. The WBST Quantitative Form contains 45 questions which help measure a broad range of mathematic skills across the three primary content domains shown in the table below.

Quantitative Skills Domain	Skills
Basic Math Computation	<ul> <li>Addition, subtraction, and division         <ul> <li>Whole numbers</li> <li>Whole monetary units</li> <li>Whole units of measure including English and metric units of time, length, weight, and distance</li> </ul> </li> </ul>
Basic Math Computation and Quantitative Evaluation	<ul> <li>Addition, subtraction, multiplication, and division         <ul> <li>Proper and improper fractions and mixed numbers</li> <li>Fractional monetary units</li> <li>Fractional units of measure including units of time and English and metric units of length, weight, and distance</li> </ul> </li> <li>Computation of rates, proportions, and percentages</li> <li>Evaluation and interpretation of line, bar or pie graphs</li> <li>Comparisons of fractional magnitudes</li> </ul>
Algebra and Geometry	<ul> <li>Computation of rates, proportions, and percentages</li> <li>Evaluation, simplification, and solving of variable expressions and equations</li> <li>Computation of length, angle, area or volume involving plane and solid geometric figures</li> <li>Understanding and use of the Pythagorean Theorem</li> </ul>

#### Significant Digits and Rounding

Precision of measurement and significant digit rules are not a basic skill that is assessed by the Test of Quantitative Skills. Therefore, all decimal fractions or measurements expressed as decimal fractions are assumed to be exact and infinitely precise. Some questions that involve the computation of percentages on units which must be expressed in whole numbers, such as people, toys, cents, plants, cats, etc., may require rounding to arrive at the correct response.

#### Item Development

In February of 1993, after an extensive review of educational and workplace skills standards, Wonderlic initiated the development of WBST test content. At the recommendation of Dr. Raju, Wonderlic commissioned a panel of language and math subject matter experts led by Hill, Katien & Associates of Wildwood, Illinois, to author field test items which would measure the occupational skills defined by the GED scale. The format of all test items was to be designed in accordance with the applied methods for teaching, assessing, and interpreting basic skills as identified in research conducted by the National Assessment of Educational Progress (NAEP), the Secretary's Commission on Achieving Necessary Skills (SCANS), and the National Education Goals Panel. This research confirmed the critical need for a new approach to basic skills assessment which is focused on job requirements and proficiency of skills application.

Under the direction of Margaret Hill and Jan Katien, a strategy for item content and presentation was developed which would achieve the stated objectives using common activities and materials familiar to the broadest possible cross-section of the population. In addition, potential issues in Differential Item Functioning (DIF) were identified to avoid the development of culture- or gender-specific test content. Finally, an initial pool of approximately 600 original test items was written which included an equal proportion of verbal and quantitative questions.

#### Field Testing

Wonderlic worked with Dr. Raju to identify the following four primary purposes for WBST field testing:

- 1. Determine the difficulty and reliability of all test items
- 2. Evaluate all test items for cultural or gender sensitivity
- 3. Identify the most appropriate time limit and number of test items for reliable final test forms
- 4. Substantiate the appropriateness of test content, design, and procedures based on the field test experience of participating schools and businesses

Six hundred and sixteen test items from the original item pool were divided across 14 field test forms - six verbal skills forms with 60 questions each and eight quantitative skills forms with 50 questions each. The first 12 items on each of these forms were common to all other verbal or quantitative field test forms, while the remaining items were unique to a given form. This provided a reliable set of anchor items which would be used to develop final test forms that provide equivalent score distributions.

Field testing was conducted between June and November of 1993. During this time, 56,000 field test forms were distributed to participating schools and businesses in a randomized serial order. This ensured that all organizations would administer an equal number of each test form and that all test items would be administered to a random cross-section of the population. At the conclusion of field testing, approximately 15,000 completed test forms were returned to Wonderlic from 126 national test sites including high schools, private vocational training schools, community colleges, and businesses. Test data collected from field testing provided over 1,000 examinee responses to each of the 616 test items included in the field testing. This data provided the foundation for constructing the final WBST forms.

#### **Test Item Characteristics**

Using both classical test analysis and item response theory on all field test data, the relative difficulty and reliability characteristics of each test item was identified. These item characteristics were statistically quantified in relation to the performance of all examinees on the common anchor items appearing on each field test form. This provided the common scale which was used to measure test item difficulty and reliability. A review of IRT test item characteristics is provided in the "Item Response Theory" section of this manual.

#### Cultural and Gender Sensitivity Review

All test items were subjected to independent expert review and statistical analysis of Differential Item Functioning (DIF) to ensure that the final test forms would be equivalent for all subgroups and free of content bias.

Under the direction of Dr. Robert G. Jones and Dr. Gowri Parameswaran of Southwest Missouri State University, a committee of graduate students reviewed all test items for cultural and gender sensitivity. A total of 118 verbal items and 52 quantitative items were identified as culture- or gender-specific. In addition, Wonderlic conducted rigorous statistical analyses of differential item functioning including the Mantel-Haenszel chi square, Lord's chi square, and Raju's area method. These analyses identified subgroup differences in test item proficiency which confirmed the findings of the independent review committee.

A team of Wonderlic research staff members evaluated each item identified by the independent review process and provided a formal written response to the committee. For the vast majority of test items, minor changes to item wording were agreed upon to provide improved clarity and balance. However, the relatively small number of remaining test items were recognized as containing subject matter that was culturally sensitive or specific. For example, a quantitative test item involving a baseball pitcher's record of balls, strikes, and innings pitched relied on knowledge specific to baseball. Such test items which contained activity or materials which are not necessarily familiar to a broad cross-section of the population were simply dropped from the item pool.

In addition, the independent review for cultural sensitivity allowed distinctions to be made between potential content bias and observed group differences in skills proficiency. As reported by the committee:

"The majority of these recommendations (38) related to questions which test knowledge of certain verb usages which were thought to be likely to be more or less familiar to people, based on cultural background. This was especially so for 1) questions in which distractors did not have personal pronouns, 2) verb forms ending in 's' or 'ed,' and 3) irregular verbs."

As indicated above, the committee had identified subject-verb agreement and the correct use of verb tenses as a likely area of subgroup differences. However, these skills are essential to one of the five basic skills identified by the SCANS commission, the ability to communicate clearly. The commission states that:

"The basic skills are the irreducible minimum for anyone who wants to get even a low skill job. They will not guarantee a career or access to a college education, but their absence will ensure that the door of opportunity remains closed." (U. S. Department of Labor, 1991)

The subject matter in the remaining test items identified by the committee as culturally sensitive involved essential job-related skills. Therefore, these items were retained in the item pool.

#### Test Length

All field test forms were administered on an extended time basis. After 15, 20, 25, and 30 minutes of testing, all examinees were instructed to code in the number of the test item on which they were working. This provided an estimate of the number of items processed by a broad range of skills groups over the timed and extended time testing periods. Field test data confirmed that most examinees will have sufficient time during a 20 minute testing period to respond to test items both at and above their skills level. Furthermore, the measurement of fluid skills application provides the most accurate assessment of an individual's proficiency in applying his or her skills in work-related situations. Therefore, the final test forms would be designed to require examinees to move quickly through continually changing test subject matter. Based on field test data and the number of test items was identified for each verbal and quantitative section of the final test forms. The verbal section would include 50 test items, while the quantitative section would include 45 test items.

Field testing provided practical experience with basic skills test administration in a wide variety of school and work environments including: public schools, vocational education programs, state social services, federal job training programs, fast food services, manufacturing and production, oil drilling, truck assembly, highway construction, and financial services. This experience substantiated the appropriateness of test content, design, and procedures.

#### **Construction of Final Test Forms**

In December of 1993, construction of WBST forms VS-1, VS-2, QS-1, and QS-2 was completed. Form equivalency was based on the test item characteristics determined in field testing. All test forms were designed based on a 20 minute timed testing period. The verbal test forms included 50 items, while the quantitative test forms included 45 items.

#### Norming of Final Test Forms

After the final test forms were constructed, a second phase of data collection was required to obtain test reliability estimates, to gather normative test data within specific education levels, occupational training programs and job titles, and to confirm the appropriateness of the 20 minute timed testing period.

Initial norming of the final WBST forms was conducted from January to June of 1994. During this time, 22,354 completed answer sheets were returned to Wonderlic from 202 national test sites including high schools, private vocational training schools, community colleges, and businesses. Submitted answer sheets included data from both timed and extended time test administrations. Additional normative data have been collected since the initial publication of the WBST in July of 1994.

For ATB purposes, the appropriate reliability estimates in normative test data for students in school and minimum score determination are presented in the "Test Reliability and Errors of Measurement," "ATB Minimum Score Determination", and "Grade Level Equivalency" sections of this manual. To confirm the appropriateness of the 20 minute timed testing period the following analyses were conducted:

#### 1. Mean Number Correct by Number Attempted

Figures 15 & 16 illustrate the average number of test items that examinees answer correctly by the number of items they attempted on both the verbal and quantitative sections of the test. These figures demonstrate that examinees at all levels of skills proficiency experience difficulty with the test content as they reach the end of the 20 minute timed testing period. In these figures, the distance between the circles and the dotted line on the graphs represent the average number of incorrect or omitted test items prior to the last item attempted. This distance remains remarkably uniform over the full range of items attempted, indicating that within the 20 minute timed testing period, all examinees are experiencing comparable difficulty relative to their skills level.

#### 2. Correlation Between Timed and Extended Time Test Scores

As discussed in the "Extended Time Test Administration" section of the WBST User's Manual for Ability-to-Benefit Testing (Wonderlic, 2011), the approved extended time WBST administration procedure combines a standard 20 minute timed period followed by a time extension on the same test form. This provides separate scores on both a timed and extended time basis. Figures 17 through 20 illustrate the increasing correlation between the number of test items answered correctly on an extended time basis and at each point during the timed testing period. These figures demonstrate that after the 20 minute timed testing period the rank order of individuals tested is stable and does not significantly change when additional testing time is given.

#### 3. Internal Consistency of Timed and Extended Time Test Administrations

In addition, as discussed in the "Test Reliability and Errors of Measurement" section of this manual, a comparative analysis of internal consistency estimates reveals that the 20 minute timed period does not practically or significantly impact internal consistency estimates for either the verbal or quantitative section of the test.

The above analyses not only confirm the appropriateness of the 20 minute time limit, but also establish that the timed testing period is sufficient to measure all primary verbal and quantitative content domains assessed by the WBST with a high degree of reliability.

Enthusiastic acceptance of the WBST by students and job applicants, confirmed support for the appropriateness of WBST content by teachers and employers, and the overwhelming consensus of participating organizations that the WBST fills an urgent need in applicant selection indicate that Wonderlic's investment in basic skills assessment is worthwhile. More importantly, a thoughtfully crafted and well executed test development plan is fundamental to achieving the content validity of a test (Nunnally & Bernstein, 1994; Nunnally, 1972; Thorndike, 1971). As recounted in this section, the WBST development process was guided by precisely this type of plan.

#### Figure 15



Source: FEK BS\R2\MANUAL\MEANCORR.PRG (08/15/96 15:09)





#### Figure 19







### ACKNOWLEDGMENTS

The Wonderlic Basic Skills Test development has been a truly collaborative effort. We need to acknowledge the very important roles played by Dr. Nambury Raju, Professor of Psychology at the Illinois Institute of Technology, and Margaret Hill and Janet Katien of Hill, Katien and Associates. Dr. Raju brought not only his preeminent expertise in item response theory (IRT), but also his extensive background in the practical realities of test development. Margaret and Jan are credited with the high quality original item pool and assisting in the test design.

Robert G. Jones, Ph.D., and Gowri Parameswaran, Ph.D., of Southwest Missouri State University and their collaborative team of graduate students provided an essential, independent review for cultural and gender sensitivity of the field test items.

#### REFERENCES

American Council on Education (1989). *Educational Measurement, (3rd ed.)*. New York, NY: Macmillan.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1985). *Standards for educational and psychological tests.* Washington, D.C.: American Psychological Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing.* Washington, DC: American Educational Research Association.

Hambleton, R.K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications.* Boston, MA: Kluwer Academic Publishers.

Hambleton, R.K., & Swaminathan, H. & Rogers, H.J. (1991). *Fundamentals of Item Response Theory.* Newbury Park, CA: Sage Publications, Inc.

National Assessment of Educational Progress (1996). NAEP 1994 *trends in academic progress.* Washington, D.C.: National Center for Education Statistics.

National Center for Education Statistics, U.S. Department of Education (1992). *The condition of education.* Washington, D.C.: Author.

Nunnally, J.C. (1972). *Educational Measurement and Evaluation (2nd ed.)*. New York, NY: McGraw-Hill, Inc.

Nunnally, J.C. & Bernstein, I.H. (1994). *Psychometric Theory*, (3rd ed.). New York, NY: McGraw-Hill, Inc.

Peterson, N.S., Kolen, M.J., & Hoover, H.D. (1989). Scaling, norming, and equating. In R.L. Linn (Ed.) *Educational measurement (3rd ed.)* (pp. 221-262). New York, NY: Macmillan.

Reckase, M.D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 4, 207-230.

Secretary's Commission on Achieving Necessary Skills, U.S. Department of Labor (1991). *What Work Requires of Schools*. Washington, D.C.: U.S. Department of Labor.

Society for Industrial and Organizational Psychology (SIOP). (2003). Principles for the Validation and Use of Personnel Selection Procedures (4th ed.). Bowling Green, OH: Author.

Thorndike, R.L. (1971). *Educational Measurement (2nd ed.)*. Washington, D.C.: American Council on Education.

U.S. Department of Education (December 1, 1995). "Student assistance general provisions; final rule. *Federal Register*, 34 CFR Part 668, Vol. 60, No. 231.

U.S. Department of Education (October 25, 1996). "List of approved ability-tobenefit tests and passing scores; notice. *Federal Register*, Vol. 61, No. 208. Washington, D.C.: U.S. Government Printing Office.

U.S. Department of Labor (1991). *Dictionary of Occupational Titles*. Washington, D.C.: Employment and Training Administration. U.S. Department of Labor.

Wonderlic, C.F., Long, E.R., & Rafilson, F. (1990). *Admissions testing at college and trade school training programs*. Libertyville, Illinois: Wonderlic, Inc.

Wonderlic, Inc. (1992). *Wonderlic Personnel Test & Scholastic Level Exam user's manual*. Libertyville, IL: Author.

Wonderlic, Inc. (2006). *Minimum Passing Scores on the Wonderlic Basic Skills Test Paper and Pencil Based Test Administrations.* Vernon Hills, IL: Author.

Wonderlic, Inc. (2011). *Wonderlic Basic Skills Test User's Manual for Ability-to-Benefit Testing.* Vernon Hills, IL: Author.