

نمونه تحلیل کیفی آزمون

دانشکده پزشکی

دانشگاه علوم پزشکی لرستان



آنالیز سوالات آزمون

آزمون: 11:08 1404/06/09										
تعداد شرکت کنندگان:		85	انحراف معیار:		0/929	میانگین ضریب تمییز:		0/21	کم	
تعداد سوال:		100	واریانس:		0/863	میانگین ضریب سختی:		0/83	آسان	
بیشترین نمره:		13/000	اعتبارسنجی KR20 :		-9/585	تعداد سوال آسان / درصد:		82 / 82	82/00%	
میانگین:		10/765	آلفای کرانباخ:		-9/585	تعداد سوال مناسب / درصد:		12 / 12	12/00%	
درصد میانگین:		82/808%	خطای استاندارد:		0/093	تعداد سوال سخت / درصد:		6 / 6	6/00%	
آزمون مرتبط:										
تعداد شرکت کنندگان:		85	انحراف معیار:		0/859	میانگین ضریب تمییز:		0/213	کم	
تعداد سوال:		100	واریانس:		0/738	میانگین ضریب سختی:		0/826	آسان	
بیشترین نمره:		13/000	اعتبارسنجی KR20 :		-11/378	تعداد سوال آسان / درصد:		82 / 82	82/000%	
میانگین:		10/723	آلفای کرانباخ:		-11/378	تعداد سوال مناسب / درصد:		12 / 12	12/000%	
درصد میانگین:		82/485%	خطای استاندارد:		0/086	تعداد سوال سخت / درصد:		6 / 6	6/000%	
Item	DISC-INDEX	RPbis	DIFF-INDEX	DIFF-INDEX	Key	Omit	A % / A	B % / B	C % / C	D % / D
1	کم	0/14	مناسب	0/44	1	1	43/53 / 37	34/12 / 29	4/71 / 4	16/47 / 14
2	کم	0/12	دشوار	0/06	4	0	10/59 / 9	83/53 / 71	0/00 / 0	5/88 / 5
3	کم	0/21	آسان	0/94	1	0	94/12 / 80	1/18 / 1	4/71 / 4	0/00 / 0
4	کم	0/00	آسان	0/95	3	0	0/00 / 0	3/53 / 3	95/29 / 81	1/18 / 1
5	کم	0/00	آسان	1/00	3	0	0/00 / 0	0/00 / 0	100/00 / 85	0/00 / 0
6	مناسب	0/55	آسان	0/98	4	0	0/00 / 0	0/00 / 0	2/35 / 2	97/65 / 83
7	کم	0/26	آسان	0/94	1	0	94/12 / 80	0/00 / 0	1/18 / 1	4/71 / 4
8	مناسب	0/42	مناسب	0/58	4	0	0/00 / 0	2/35 / 2	40/00 / 34	57/65 / 49
9	کم	0/00	آسان	0/80	1	0	80/00 / 68	16/47 / 14	3/53 / 3	0/00 / 0
10	کم	0/00	آسان	1/00	3	0	0/00 / 0	0/00 / 0	100/00 / 85	0/00 / 0
11	مناسب	0/38	آسان	0/93	1	0	92/94 / 79	0/00 / 0	4/71 / 4	2/35 / 2
12	کم	0/24	آسان	0/95	1	0	95/29 / 81	1/18 / 1	2/35 / 2	1/18 / 1
13	کم	0/21	آسان	0/98	3	0	0/00 / 0	1/18 / 1	97/65 / 83	1/18 / 1



آنالیز سوالات آزمون

آزمون مرتبط:										
تعداد شرکت کنندگان:	85	انحراف معیار:	0/859	میانگین ضریب تمیز:	0/213	کم				
تعداد سوال:	100	واریانس:	0/738	میانگین ضریب سختی:	0/826	آسان				
بیشترین نمره:	13/000	اعتبارسنجی KR20 :	-11/378	تعداد سوال آسان / درصد:	82 / 82/000%					
میانگین:	10/723	آلفای کرانباخ:	-11/378	تعداد سوال مناسب / درصد:	12 / 12/000%					
درصد میانگین:	82/485%	خطای استاندارد:	0/086	تعداد سوال سخت / درصد:	6 / 6/000%					
Item	DISC-INDEX	RPbis	DIFF-INDEX	DIFF-INDEX	Key	Omit	A% / A	B% / B	C% / C	D% / D
14	مناسب	0/42	آسان	0/94	1	0	94/12 / 80	3/53 / 3	2/35 / 2	0/00 / 0
15	مناسب	0/36	مناسب	0/64	3	0	5/88 / 5	12/94 / 11	63/53 / 54	17/65 / 15
16	کم	0/21	آسان	0/98	2	0	1/18 / 1	97/65 / 83	1/18 / 1	0/00 / 0
17	کم	0/09	آسان	0/97	3	0	3/53 / 3	0/00 / 0	96/47 / 82	0/00 / 0
18	کم	0/00	آسان	1/00	3	0	0/00 / 0	0/00 / 0	100/00 / 85	0/00 / 0
19	مناسب	0/45	آسان	0/99	2	0	1/18 / 1	98/82 / 84	0/00 / 0	0/00 / 0
20	مناسب	0/45	آسان	0/89	4	0	0/00 / 0	3/53 / 3	7/06 / 6	89/41 / 76
21	مناسب	0/30	آسان	0/95	1	0	95/29 / 81	1/18 / 1	3/53 / 3	0/00 / 0
22	مناسب	0/31	آسان	0/94	2	0	2/35 / 2	94/12 / 80	1/18 / 1	2/35 / 2
23	کم	0/22	آسان	0/86	3	0	10/59 / 9	3/53 / 3	85/88 / 73	0/00 / 0
24	کم	0/21	آسان	0/99	2	0	1/18 / 1	98/82 / 84	0/00 / 0	0/00 / 0
25	کم	0/23	آسان	0/97	1	0	96/47 / 82	1/18 / 1	0/00 / 0	2/35 / 2
26	کم	0/11	مناسب	0/61	3	0	23/53 / 20	8/24 / 7	61/18 / 52	7/06 / 6
27	مناسب	0/44	آسان	0/91	2	0	5/88 / 5	90/59 / 77	2/35 / 2	1/18 / 1
28	کم	0/16	مناسب	0/55	1	1	55/29 / 47	11/77 / 10	20/00 / 17	11/77 / 10
29	مناسب	0/53	آسان	0/94	3	0	0/00 / 0	3/53 / 3	94/12 / 80	2/35 / 2
30	مناسب	0/37	آسان	0/94	1	0	94/12 / 80	2/35 / 2	0/00 / 0	3/53 / 3
31	کم	0/05	آسان	0/98	3	0	0/00 / 0	2/35 / 2	97/65 / 83	0/00 / 0
32	کم	0/00	آسان	1/00	4	0	0/00 / 0	0/00 / 0	0/00 / 0	100/00 / 85
33	مناسب	0/30	آسان	0/84	4	0	2/35 / 2	0/00 / 0	14/12 / 12	83/53 / 71



دانشگاه علوم پزشکی لرستان
و خدمات بهداشتی درمانی لرستان

دانشگاه علوم پزشکی لرستان

آنالیز سوالات آزمون

آزمون مرتبط:										
تعداد شرکت کنندگان:		85	انحراف معیار:		0/859	میانگین ضریب تمیز:		0/213	کم	
تعداد سوال:		100	واریانس:		0/738	میانگین ضریب سختی:		0/826	آسان	
بیشترین نمره:		13/000	اعتبارسنجی KR20 :		-11/378	تعداد سوال آسان / درصد:		82	82/000% /	
میانگین:		10/723	آلفای کرانباخ:		-11/378	تعداد سوال مناسب / درصد:		12	12/000% /	
درصد میانگین:		82/485%	خطای استاندارد:		0/086	تعداد سوال سخت / درصد:		6	6/000% /	
Item	DISC-INDEX	RPbis	DIFF-INDEX	DIFF-INDEX	Key	Omit	A% / A	B% / B	C% / C	D% / D
34	کم	0/29	آسان	0/81	2	1	3/53 / 3	81/18 / 69	7/06 / 6	7/06 / 6
35	کم	0/11	آسان	0/81	2	0	2/35 / 2	81/18 / 69	1/18 / 1	15/29 / 13
36	مناسب	0/40	آسان	0/74	1	0	74/12 / 63	20/00 / 17	2/35 / 2	3/53 / 3
37	مناسب	0/43	آسان	0/87	2	0	7/06 / 6	87/06 / 74	3/53 / 3	2/35 / 2
38	مناسب	0/31	آسان	0/94	1	0	94/12 / 80	1/18 / 1	3/53 / 3	1/18 / 1
39	مناسب	0/40	آسان	0/91	2	0	8/24 / 7	90/59 / 77	0/00 / 0	1/18 / 1
40	کم	0/24	آسان	0/95	2	0	0/00 / 0	95/29 / 81	2/35 / 2	2/35 / 2
41	مناسب	0/42	آسان	0/73	1	0	72/94 / 62	21/18 / 18	1/18 / 1	4/71 / 4
42	کم	0/22	آسان	0/80	3	0	1/18 / 1	7/06 / 6	80/00 / 68	11/77 / 10
43	کم	0/16	آسان	0/97	1	0	96/47 / 82	2/35 / 2	1/18 / 1	0/00 / 0
44	کم	0/15	دشوار	0/25	2	0	35/29 / 30	24/71 / 21	23/53 / 20	16/47 / 14
45	کم	0/21	آسان	0/99	2	0	1/18 / 1	98/82 / 84	0/00 / 0	0/00 / 0
46	کم	0/00	آسان	1/00	3	0	0/00 / 0	0/00 / 0	100/00 / 85	0/00 / 0
47	کم	0/13	آسان	0/98	1	0	97/65 / 83	0/00 / 0	2/35 / 2	0/00 / 0
48	کم	0/22	آسان	0/88	1	1	88/24 / 75	1/18 / 1	4/71 / 4	4/71 / 4
49	کم	0/09	آسان	0/97	4	0	1/18 / 1	2/35 / 2	0/00 / 0	96/47 / 82
50	کم	0/20	آسان	0/89	2	0	1/18 / 1	89/41 / 76	1/18 / 1	8/24 / 7
51	کم	0/00	آسان	1/00	2	0	0/00 / 0	100/00 / 85	0/00 / 0	0/00 / 0
52	کم	0/06	آسان	0/95	4	0	1/18 / 1	3/53 / 3	0/00 / 0	95/29 / 81
53	مناسب	0/31	مناسب	0/51	2	0	40/00 / 34	50/59 / 43	5/88 / 5	3/53 / 3



دانشگاه علوم پزشکی لرستان

آنالیز سوالات آزمون

آزمون مرتبط:										
تعداد شرکت کنندگان:		85	انحراف معیار:		0/859	میانگین ضریب تمیز:		0/213	کم	
تعداد سوال:		100	واریانس:		0/738	میانگین ضریب سختی:		0/826	آسان	
بیشترین نمره:		13/000	اعتبارسنجی KR20 :		-11/378	تعداد سوال آسان / درصد:		82	82/000% /	
میانگین:		10/723	آلفای کرانباخ:		-11/378	تعداد سوال مناسب / درصد:		12	12/000% /	
درصد میانگین:		82/485%	خطای استاندارد:		0/086	تعداد سوال سخت / درصد:		6	6/000% /	
Item	DISC-INDEX	RPbis	DIFF-INDEX	DIFF-INDEX	Key	Omit	A % / A	B % / B	C % / C	D % / D
54	کم	0/22	آسان	0/91	3	0	1/18 / 1	1/18 / 1	90/59 / 77	7/06 / 6
55	مناسب	0/38	آسان	0/88	3	0	2/35 / 2	7/06 / 6	88/24 / 75	2/35 / 2
56	کم	0/10	آسان	0/88	4	0	7/06 / 6	2/35 / 2	2/35 / 2	88/24 / 75
57	مناسب	0/35	مناسب	0/66	1	0	65/88 / 56	22/35 / 19	10/59 / 9	1/18 / 1
58	مناسب	0/38	آسان	0/98	4	0	0/00 / 0	0/00 / 0	2/35 / 2	97/65 / 83
59	کم	0/20	مناسب	0/32	4	0	55/29 / 47	9/41 / 8	3/53 / 3	31/77 / 27
60	کم	0/09	آسان	0/79	3	0	4/71 / 4	1/18 / 1	78/82 / 67	15/29 / 13
61	مناسب	0/50	آسان	0/84	2	0	11/77 / 10	83/53 / 71	2/35 / 2	2/35 / 2
62	کم	0/09	آسان	0/91	1	0	90/59 / 77	1/18 / 1	4/71 / 4	3/53 / 3
63	کم	0/18	آسان	0/95	4	0	3/53 / 3	0/00 / 0	1/18 / 1	95/29 / 81
64	مناسب	0/30	آسان	0/88	3	0	0/00 / 0	1/18 / 1	88/24 / 75	10/59 / 9
65	کم	0/20	آسان	0/92	3	0	3/53 / 3	3/53 / 3	91/77 / 78	1/18 / 1
66	کم	0/08	مناسب	0/65	3	0	0/00 / 0	0/00 / 0	64/71 / 55	35/29 / 30
67	مناسب	0/40	آسان	0/91	4	0	7/06 / 6	1/18 / 1	1/18 / 1	90/59 / 77
68	مناسب	0/37	آسان	0/77	2	0	20/00 / 17	76/47 / 65	1/18 / 1	2/35 / 2
69	مناسب	0/38	آسان	0/93	2	0	2/35 / 2	92/94 / 79	3/53 / 3	1/18 / 1
70	کم	0/00	دشوار	0/00	3	0	8/24 / 7	10/59 / 9	0/00 / 0	81/18 / 69
71	کم	0/00	آسان	1/00	3	0	0/00 / 0	0/00 / 0	100/00 / 85	0/00 / 0
72	کم	0/07	دشوار	0/15	2	0	0/00 / 0	15/29 / 13	0/00 / 0	84/71 / 72
73	کم	0/22	آسان	0/91	4	0	5/88 / 5	2/35 / 2	1/18 / 1	90/59 / 77



دانشگاه علوم پزشکی
و خدمات بهداشتی درمانی لرستان

دانشگاه علوم پزشکی لرستان

آنالیز سوالات آزمون

آزمون مرتبط:										
تعداد شرکت کنندگان:		85	انحراف معیار:		0/859	میانگین ضریب تمیز:		0/213	کم	
تعداد سوال:		100	واریانس:		0/738	میانگین ضریب سختی:		0/826	آسان	
بیشترین نمره:		13/000	اعتبارسنجی KR20 :		-11/378	تعداد سوال آسان / درصد:		82 / 82/000%		
میانگین:		10/723	آلفای کرانباخ:		-11/378	تعداد سوال مناسب / درصد:		12 / 12/000%		
درصد میانگین:		82/485%	خطای استاندارد:		0/086	تعداد سوال سخت / درصد:		6 / 6/000%		
Item	DISC-INDEX	RPbis	DIFF-INDEX	DIFF-INDEX	Key	Omit	A % / A	B % / B	C % / C	D % / D
74	مناسب	0/47	آسان	0/87	4	0	0/00 / 0	9/41 / 8	3/53 / 3	87/06 / 74
75	مناسب	0/44	آسان	0/86	1	0	85/88 / 73	11/77 / 10	0/00 / 0	2/35 / 2
76	کم	0/23	آسان	0/71	3	0	5/88 / 5	12/94 / 11	70/59 / 60	10/59 / 9
77	کم	0/05	مناسب	0/68	2	0	20/00 / 17	68/24 / 58	11/77 / 10	0/00 / 0
78	کم	0/00	آسان	1/00	4	0	0/00 / 0	0/00 / 0	0/00 / 0	100/00 / 85
79	مناسب	0/47	آسان	0/98	3	0	0/00 / 0	1/18 / 1	97/65 / 83	1/18 / 1
80	نامناسب	-0/42	دشوار	0/15	2	0	5/88 / 5	15/29 / 13	0/00 / 0	78/82 / 67
81	کم	0/00	آسان	1/00	4	0	0/00 / 0	0/00 / 0	0/00 / 0	100/00 / 85
82	مناسب	0/38	آسان	0/78	4	0	0/00 / 0	20/00 / 17	2/35 / 2	77/65 / 66
83	کم	0/24	آسان	0/87	1	0	87/06 / 74	1/18 / 1	4/71 / 4	7/06 / 6
84	مناسب	0/42	آسان	0/77	1	1	76/47 / 65	3/53 / 3	8/24 / 7	10/59 / 9
85	مناسب	0/36	آسان	0/97	2	0	0/00 / 0	96/47 / 82	3/53 / 3	0/00 / 0
86	کم	0/10	آسان	0/94	4	0	4/71 / 4	0/00 / 0	1/18 / 1	94/12 / 80
87	کم	0/15	آسان	0/92	1	0	91/77 / 78	5/88 / 5	0/00 / 0	2/35 / 2
88	نامناسب	-0/03	آسان	0/99	2	0	1/18 / 1	98/82 / 84	0/00 / 0	0/00 / 0
89	مناسب	0/45	آسان	0/99	1	0	98/82 / 84	0/00 / 0	1/18 / 1	0/00 / 0
90	کم	0/20	آسان	0/92	4	0	2/35 / 2	1/18 / 1	4/71 / 4	91/77 / 78
91	کم	0/00	آسان	0/95	1	0	95/29 / 81	1/18 / 1	0/00 / 0	3/53 / 3
92	کم	0/09	آسان	0/81	4	1	1/18 / 1	11/77 / 10	4/71 / 4	81/18 / 69
93	نامناسب	-0/08	آسان	0/92	2	0	2/35 / 2	91/77 / 78	2/35 / 2	3/53 / 3



آنالیز سوالات آزمون

آزمون مرتبط:										
تعداد شرکت کنندگان:		85	انحراف معیار:		0/859	میانگین ضریب تمیز:		0/213	کم	
تعداد سوال:		100	واریانس:		0/738	میانگین ضریب سختی:		0/826	آسان	
بیشترین نمره:		13/000	اعتبارسنجی KR20 :		-11/378	تعداد سوال آسان / درصد:		82 / 82/000%		
میانگین:		10/723	آلفای کرانباخ:		-11/378	تعداد سوال مناسب / درصد:		12 / 12/000%		
درصد میانگین:		82/485%	خطای استاندارد:		0/086	تعداد سوال سخت / درصد:		6 / 6/000%		
Item	DISC-INDEX	RPbis	DIFF-INDEX	DIFF-INDEX	Key	Omit	A% / A	B% / B	C% / C	D% / D
94	کم	0/08	مناسب	0/51	4	0	14/12 / 12	4/71 / 4	30/59 / 26	50/59 / 43
95	کم	0/25	آسان	0/92	1	0	91/77 / 78	1/18 / 1	3/53 / 3	3/53 / 3
96	کم	0/00	آسان	0/95	4	0	3/53 / 3	0/00 / 0	1/18 / 1	95/29 / 81
97	کم	0/15	دشوار	0/01	4	1	44/71 / 38	48/24 / 41	4/71 / 4	1/18 / 1
98	مناسب	0/31	مناسب	0/49	1	1	49/41 / 42	20/00 / 17	4/71 / 4	24/71 / 21
99	کم	0/09	آسان	0/99	4	0	1/18 / 1	0/00 / 0	0/00 / 0	98/82 / 84
100	مناسب	0/33	آسان	0/99	3	0	0/00 / 0	1/18 / 1	98/82 / 84	0/00 / 0

گزارش تحلیلی آزمون به سه شکل: ۱- جدول، ۲. متن مختصر و ۳. متن مفصل ارائه می شود:

۱- جدول گزارش مختصر تحلیل شاخص های آزمون

شاخص	تعریف	مقدار آزمون	استاندارد مطلوب	تفسیر و مثال
میانگین نمرات	میانگین عملکرد کل داوطلبان.	۱۰.۷۶ از ۱۳ (۸۲.۸٪)	۵۰-۷۰٪	بالا ← آزمون آسان.
درصد میانگین	نسبت میانگین به نمره کل در درصد.	۸۲.۸٪	۵۰-۷۰٪	آزمون آسان تر از حد مطلوب.
انحراف معیار (SD)	پراکندگی نمرات.	۰.۹۲۹	۲-۳ (برای آزمون ۱۳ نمره ای)	پایین ← نمرات شبیه هم.
واریانس	شاخص پراکندگی (مجذور SD).	۰.۸۶۳	—	تأیید SD پایین.
KR-20	شاخص پایایی آزمون های چندگزینه ای.	۰.۵۸۵	≥ 0.7	منفی ← پایایی صفر. غیر قابل قبول
آلفای کرونباخ	همابستگی درونی آیتم ها.	۰.۵۸۵	≥ 0.7	منفی ← آزمون معیوب. غیر قابل قبول
SEM	خطای استاندارد اندازه گیری.	۰.۰۹۳	کمتر، بهتر	به ظاهر خوب اما به دلیل پایایی منفی بی اعتبار.
DISC-INDEX	قدرت تفکیک سؤال بین قوی و ضعیف.	۰.۲۱ میانگین	≥ 0.3 خوب	پایین ← بیشتر سؤالات قدرت تفکیک ندارند. مثال: سؤال ۴۴ با ۰.۱۵ ضعیف است.
RPbis	همبستگی پاسخ درست هر سؤال با کل نمره.	۰.۱-۰.۴ اغلب	≥ 0.3 خوب	در برخی سؤالات منفی ← سؤال معیوب (مثل سؤال ۸۰).
DIFF-INDEX	نسبت پاسخ صحیح به سؤال.	۰.۸۳ میانگین	۰.۳-۰.۷	بسیار آسان (مثال: سؤال ۵ با ۱.۰۰).
Omit	درصد پاسخ ندادن به سؤال.	در برخی سؤالات بالای ۵٪	$> 5\%$	برخی سؤالات مبهم/سخت.
توزیع دشواری	درصد سؤالات آسان/متوسط/سخت.	آسان ۸۲٪، متوسط ۱۲٪، سخت ۶٪	۳۰٪ / ۴۰٪ / ۳۰٪	تعادل رعایت نشده.
Distractors	گزینه های غلط باید جذاب باشند.	اغلب ناکارآمد	هر گزینه توسط ۵ تا ۱۰ درصد شرکت کنندگان انتخاب شود	بسیاری از گزینه ها اصلاً انتخاب نشده اند ← بی اثر.

۲- گزارش مختصر تحلیل آزمون

۱. میانگین نمرات (Mean Score)

- تعریف: میانگین نشان‌دهنده سطح کلی عملکرد شرکت‌کنندگان در آزمون است و از تقسیم مجموع نمرات به تعداد شرکت‌کنندگان به‌دست می‌آید.
- نتیجه آزمون: ۱۰.۷۶ از ۱۳ (۸۲.۸٪).
- استاندارد: بهتر است میانگین حدود ۷۰-۵۰٪ باشد تا آزمون قدرت تمایز داشته باشد.
- تفسیر: بالا بودن میانگین یعنی آزمون آسان بوده است.
- مثال: اگر در یک کلاس ۲۰ نفره، ۱۸ نفر نمره بالای ۱۸ بگیرند، میانگین زیاد است ولی تمایز ایجاد نمی‌شود.

۲. درصد میانگین (Mean Percentage)

- تعریف: نسبت میانگین به حداکثر نمره در مقیاس درصد.
- نتیجه آزمون: ۸۲.۸٪.
- استاندارد: ۷۰-۵۰٪ مطلوب.
- تفسیر: آزمون آسان‌تر از حد استاندارد بوده است.

۳. انحراف معیار (Standard Deviation, SD)

- تعریف: پراکندگی نمرات نسبت به میانگین؛ هرچه بیشتر باشد، تنوع نمرات بیشتر است.
- نتیجه آزمون: ۰.۹۲۹.
- استاندارد: حدود ۲۰-۱۰٪ نمره کل (در این مورد باید بین ۱.۳ تا ۲.۶ باشد).
- تفسیر: نمرات تقریباً شبیه هم بوده‌اند پس آزمون افتراق ایجاد نکرده.
- مثال: اگر همه بین ۱۹ و ۲۰ بگیرند، SD پایین خواهد بود.

۴. واریانس (Variance)

- تعریف: مجذور انحراف معیار و نشان‌دهنده پراکندگی نمرات است.
- نتیجه آزمون: ۰.۸۶۳.
- تفسیر: پایین است و تأیید می‌کند که نمرات شبیه هم هستند.

۵. اعتبارسنجی (KR-20 (Kuder-Richardson Formula 20)

- تعریف: شاخص پایایی آزمون‌های چندگزینه‌ای (۰ تا ۱).
- نتیجه آزمون: ۰.۵۸۵- (غیرمنطقی و منفی).
- استاندارد: بالاتر از ۰.۷ خوب؛ کمتر از ۰.۵ ضعیف.
- تفسیر: آزمون پایایی ندارد، احتمال وجود خطا در طراحی سؤالات یا کلید.
- مثال: اگر در یک آزمون، بیشتر گزینه‌ها اشتباه طراحی شده باشند، KR-20 منفی می‌شود.

۶. آلفای کرونباخ (Cronbach's Alpha)

- تعریف: شاخص پایایی و هماهنگی درونی آیتم‌ها (۰ تا ۱).
- نتیجه آزمون: ۰.۵۸۵- (غیرعادی).
- استاندارد: ≥ 0.7 مطلوب.
- تفسیر: آزمون پایایی بسیار ضعیف دارد.

۷. خطای استاندارد اندازه‌گیری (SEM)

- تعریف: دقت اندازه‌گیری آزمون را نشان می‌دهد؛ فاصله‌ای که نمره واقعی فرد ممکن است با نمره مشاهده شده تفاوت داشته باشد.
- نتیجه آزمون: ۰.۰۹۳.
- استاندارد: هرچه کمتر، بهتر.
- تفسیر: به ظاهر خوب است، اما چون شاخص پایایی منفی است، این شاخص معتبر نیست.
- مثال: اگر دانشجویی ۱۰ از ۲۰ بگیرد، نمره واقعی او بین ۹.۹ تا ۱۰.۱ است.

۸. ضریب تمیز (Discrimination Index)

- تعریف: توانایی سؤال در تمایز بین دانشجویان قوی و ضعیف.
- نتیجه آزمون: میانگین ۰.۲۱.
- استاندارد: ≥ 0.3 خوب، $0.2 - 0.29$ قابل قبول، < 0.2 ضعیف.
- تفسیر: اکثر سؤالات تمایز کافی ندارند.
- مثال: اگر هم دانشجویان قوی و هم ضعیف یک سؤال را درست بزنند، ضریب تمیز آن کم است.

۹. ضریب دشواری (Difficulty Index)

- تعریف: نسبت پاسخ‌دهندگان صحیح به کل پاسخ‌دهندگان.
- نتیجه آزمون: میانگین ۰.۸۳.
- استاندارد: ۰.۷-۰.۳ مناسب.
- تفسیر: آزمون به‌طور کلی خیلی آسان بوده است.

۱۰. توزیع دشواری سؤالات

- آسان: ۸۲٪
- متوسط: ۱۲٪
- سخت: ۶٪
- استاندارد: ۳۰٪ آسان، ۴۰٪ متوسط، ۳۰٪ سخت.
- تفسیر: آزمون تعادل ندارد و شدیداً به سمت آسانی رفته است.

۱۱. همبستگی نقطه‌ای-دو وجهی یا دو رشته‌ای (Point Biserial Correlation – R_{Pbis})

- تعریف: همبستگی بین پاسخ صحیح به یک سؤال (دو وجهی: درست/غلط) و کل نمره آزمون.
- محدوده: -۱ تا +۱.

- استاندارد: ≥ 0.3 خوب، $0.2-0.29$ مرزی، < 0.2 ضعیف.
 - کاربرد: مشابه ضریب تمیز ولی دقیق‌تر و مبتنی بر همبستگی.
 - مثال: اگر RPbis برای سؤال ۸۰ منفی باشد، یعنی دانشجویان قوی بیشتر آن را غلط جواب داده‌اند، پس سؤال معیوب است.
-

۱۲. گزینه‌های انتخاب نشده (Omit)

- تعریف: درصد داوطلبانی که به سؤال پاسخ نداده‌اند.
 - استاندارد: معمولاً باید کمتر از ۵٪ باشد؛ بالاتر از آن نشانه ابهام در سؤال یا دشواری بیش از حد است.
 - مثال: اگر در سؤال ۷۰، ۱۲٪ شرکت‌کنندگان هیچ گزینه‌ای انتخاب نکرده باشند، آن سؤال نیاز به بازنگری دارد.
-

۱۳. گزینه‌های انحرافی (Distractors)

- تعریف: گزینه‌های غلطی که باید بخشی از پاسخ‌دهندگان را جذب کنند.
 - وضعیت آزمون: بسیاری از گزینه‌ها اصلاً انتخاب نشده‌اند.
 - استاندارد: هر گزینه باید حداقل ۱۰-۵٪ دانشجویان را به خود جذب کند.
 - تفسیر: گزینه‌های انحرافی ناکارآمدند پس سؤال‌ها کیفیت پایین دارند.
 - مثال: اگر سؤال ۴ گزینه‌ای باشد و ۹۵٪ دانشجویان فقط گزینه صحیح را انتخاب کنند، سایر گزینه‌ها بی‌اثرند.
-

۳- گزارش مفصل تحلیل روان‌سنجی آزمون

تعاریف و توضیح شاخص‌های روان‌سنجی آزمون‌ها

این شاخص‌ها به دو دسته اصلی تقسیم می‌شوند: شاخص‌های توصیفی (ویژگی‌های کلی آزمون) و شاخص‌های تحلیل سؤال (Item)

(Analysis)

الف) شاخص‌های توصیفی کلی آزمون

۱. تعداد شرکت‌کنندگان (N)

• توضیح: حجم نمونه برگزار کننده آزمون.

• تفسیر: هرچه تعداد بیشتر باشد، نتایج پایدارتر و قابل تعمیم‌تر است. معمولاً حداقل ۳۰ نفر برای تحلیل قابل قبول است.

۲. میانگین (Mean)

• توضیح: متوسط نمرات کسب شده توسط شرکت‌کنندگان.

• تفسیر: نشان‌دهنده مرکزیت داده‌هاست. میانگین بسیار بالا ممکن است نشان‌دهنده آسانی آزمون و میانگین بسیار پایین نشان‌دهنده دشواری بیش از حد آن باشد.

۳. درصد میانگین (Mean Percentage)

• توضیح: میانگین نمرات به صورت درصدی از نمره کل.

• تفسیر: درک آن آسان‌تر است. معمولاً درصد بین ۵۰٪ تا ۷۰٪ مطلوب در نظر گرفته می‌شود.

۴. انحراف معیار (Standard Deviation - SD)

• توضیح: نشان‌دهنده میزان پراکندگی یا variability نمرات حول میانگین است.

• تفسیر:

SD *پایین (مثلاً زیر ۱): نمرات به هم فشرده هستند. آزمون نتوانسته است بین افراد قوی و ضعیف تمایز قائل شود.

SD *بالا: نمرات پراکنده هستند. آزمون به خوبی افراد را از هم متمایز کرده است. SD مطلوب معمولاً بین ۱.۵ تا ۲.۵ (برای نمره کل ۱۳) است.

۵. واریانس (Variance)

• توضیح: مربع انحراف معیار. همان مفهوم پراکندگی را اندازه‌گیری می‌کند.

• تفسیر: تفسیر آن مشابه انحراف معیار است.

۶. پایایی (Reliability)

• توضیح: مهمترین شاخص. نشان‌دهنده ثبات، پایایی و یکنواختی نتایج آزمون است. آیا اگر آزمون را دوباره برگزار کنیم، نتایج مشابه خواهد بود؟

• انواع شاخص‌های پایایی و تفسیر آنها:

*ضریب آلفای کرونباخ (Cronbach's Alpha) و اعتبارسنجی KR-20: رایج‌ترین شاخص‌ها برای آزمون‌های چندگزینه‌ای.

○ مقدار $0.9 \leq \alpha$: عالی (عالی برای آزمون‌های حساس مانند کنکور)

○ مقدار $0.8 \leq \alpha < 0.9$: خوب (برای اکثر آزمون‌های استاندارد)

○ مقدار $0.7 \leq \alpha < 0.8$: قابل قبول

○ مقدار $0.6 \leq \alpha < 0.7$: مشکوک (نیاز به بازبینی دارد)

○ مقدار $\alpha < 0.6$: غیرقابل قبول

○ مقدار منفی: یک خطای فاحش. نشان دهنده ناهماهنگی شدید بین سؤالات، احتمالاً کلید سؤالات نادرست بوده است.

۷. خطای استاندارد اندازه‌گیری (SEM - Standard Error of Measurement)

- توضیح: تخمینی از میزان خطایی که در نمره هر فرد وجود دارد. به زبان ساده، یک "بازه اطمینان" حول نمره واقعی فرد است.
- تفسیر: هر چه SEM کوچک‌تر باشد، دقت اندازه‌گیری بیشتر است. از رابطه $SEM = SD * \sqrt{1 - Reliability}$ به دست می‌آید.

ب) شاخص‌های تحلیل سؤال (Item Analysis)

۸. ضریب دشواری (p - Difficulty Index)

- توضیح: نسبت افرادی که به سؤال پاسخ صحیح داده‌اند.

$$p = \frac{\text{تعداد کل شرکت‌کنندگان}}{\text{تعداد پاسخ صحیح}}$$

- تفسیر:

$$p > 0.75: \text{سؤال آسان}$$

$$0.4 \leq p \leq 0.75: \text{سؤال مناسب (دشواری متوسط)}$$

$$p < 0.4: \text{سؤال دشوار}$$

- هدف: داشتن ترکیبی از سؤالات آسان، متوسط و دشوار. معمولاً توزیع ۲۰٪ آسان، ۶۰٪ متوسط، ۲۰٪ دشوار مطلوب است.

۹. ضریب تمییز (D - Discrimination Index)

- توضیح: نشان می‌دهد یک سؤال تا چه حد بین افراد قوی و ضعیف (با توجه نمره کل هر فرد) تمایز قائل می‌شود.
- محاسبه: شرکت‌کنندگان را بر اساس نمره کل به دو گروه قوی (۲۷٪ بالا) و ضعیف (۲۷٪ پایین) تقسیم می‌کنیم.

$$D = \frac{\text{تعداد افراد در هر گروه}}{\text{تعداد پاسخ صحیح در گروه قوی} - \text{تعداد پاسخ صحیح در گروه ضعیف}}$$

- تفسیر:

$$D \geq 0.4: \text{سؤال عالی (به خوبی تمایز ایجاد می‌کند)}$$

$$0.3 \leq D < 0.4: \text{سؤال خوب (قابل قبول، ممکن است نیاز به کمی بهبود داشته باشد)}$$

$$0.2 \leq D < 0.3: \text{سؤال ضعیف (نیاز به بازبینی جدی دارد)}$$

$$D < 0.2: \text{سؤال غیرقابل قبول (باید حذف یا کاملاً بازنویسی شود)}$$

D منفی: یک فاجعه، نشان می‌دهد افراد ضعیف بیشتر از افراد قوی به این سؤال پاسخ صحیح داده‌اند. احتمال کلید اشتباه یا

ابهام شدید در سؤال.

۱۰. همبستگی نقطه‌ای دو رشته‌ای (RPbis - Point-Biserial Correlation)

- توضیح: یک شاخص پیشرفته‌تر برای تمایز. همبستگی بین پاسخ به یک سؤال خاص (صحیح=۱، غلط=۰) و نمره کل آزمون را اندازه می‌گیرد.
- تفسیر: مشابه ضریب تمییز است. مقادیر مثبت و بالا (بالای ۰.۳) مطلوب هستند. مقادیر نزدیک به صفر یا منفی نشان‌دهنده عملکرد ضعیف سؤال است.

۱۱. تحلیل گزینه‌های انحرافی (Distractor Analysis)

- توضیح: نشان می‌دهد که هر گزینه غلط (انحرافی) تا چه حد توسط شرکت‌کنندگان، به‌ویژه گروه ضعیف، انتخاب شده است.
- تفسیر: یک گزینه انحرافی خوب باید:

*توسط تعدادی از شرکت‌کنندگان (حداقل ۵٪) انتخاب شود.

*در گروه ضعیف بیشتر از گروه قوی انتخاب شود.

*اگر گزینه‌ای توسط هیچکس انتخاب نشود، بی‌فایده است و باید جایگزین شود. اگر یک گزینه انحرافی توسط افراد قوی بیشتر انتخاب شود، احتمالاً گزینه همراه‌کننده یا غلط است.

تحلیل مفصل شاخص‌های روانسنجی آزمون و پیشنهادات

۱. میانگین و درصد میانگین

- مقدار: ۱۰.۷۶ از ۱۳ معادل ۸۲.۸٪
- تفسیر: میانگین بالا نشان‌دهنده عملکرد خوب دانشجویان است، اما ممکن است نشان‌دهنده آسانی بیش از حد آزمون نیز باشد.
- مقایسه با استاندارد: میانگین ایده‌آل معمولاً بین ۷۰-۶۰ درصد است.
- پیشنهاد: افزایش تعادل در سطح دشواری سؤالات.

۲. انحراف معیار و واریانس

- انحراف معیار: ۰.۹۲۹ (پایین)
- واریانس: ۰.۸۶۳ (پایین)
- تفسیر: نشان‌دهنده همگنی زیاد نمرات و عدم تمایز کافی بین دانشجویان قوی و ضعیف.
- پیشنهاد: افزودن سؤالات با دشواری متوسط و دشوار به منظور افزایش پراکندگی نمرات.

وضعیت آزمون از لحاظ شاخص‌های مرکزی و پراکندگی

شاخص	مقدار	تفسیر و وضعیت
میانگین نمره	۱۰.۷۲ از ۱۳ (۸۲.۴۸۵٪)	قابل قبول. نشان‌دهنده تسلط نسبی اکثر دانشجویان بر محتوای آموزشی است.
انحراف معیار	۰.۹۲۹	پایین. نشان‌دهنده همگنی بالا و تمرکز نمرات حول میانگین است. این امر توانایی آزمون در تفکیک دانشجویان را کاهش می‌دهد.
واریانس	۰.۸۶۳	پایین. تأییدکننده همسانی و عدم پراکندگی کافی در نمرات است.

۳. ضریب دشواری (Difficulty Index)

- میانگین: ۰.۸۳ (آسان)
- تفسیر: آزمون بسیار آسان است. تنها ۶٪ سؤالات دشوار هستند.
- مثال: سؤال ۵ با ضریب دشواری ۱.۰۰ (همه پاسخ صحیح دادند) نمونه‌ای از سؤال بسیار آسان است.
- پیشنهاد: کاهش سؤالات بسیار آسان و افزایش سؤالات با دشواری متوسط (۰.۴-۰.۶).

جدول شاخص دشواری سؤالات (Difficulty Index)

سطح دشواری	تعداد سؤال	درصد	تفسیر و وضعیت
آسان (بیش از ۰.۷۵)	۸۲	۸۲٪	بیش از حد آسان. این حجم از سؤالات آسان باعث می‌شود آزمون نتواند به خوبی دانشجویان متوسط و قوی را از یکدیگر متمایز کند.
مناسب (۰.۴ تا ۰.۷۵)	۱۲	۱۲٪	کم بود. تعداد سؤالات با سطح دشواری متوسط که برای تمایزگذاری ایده‌آل هستند، ناکافی است.
دشوار (کمتر از ۰.۴)	۶	۶٪	کم بود. وجود سؤالات چالشی برای شناسایی دانشجویان برتر ضروری است.

میانگین ضریب دشواری: ۰.۸۲۶ (آسان)

نتیجه‌گیری: عدم تعادل آشکار در سطح دشواری سؤالات، اصلی‌ترین ضعف آزمون حاضر است

۴. ضریب تمییز (Discrimination Index)

- میانگین: ۰.۲۱ (پایین)
- تفسیر: سؤالات به خوبی بین دانشجویان قوی و ضعیف تمایز قائل نمی‌شوند.
- مثال: سؤال ۸۰ با ضریب تمییز منفی (۰.۴۲-) نشان می‌دهد که حتی دانشجویان ضعیف بهتر پاسخ داده‌اند.
- پیشنهاد: بازبینی سؤالات با تمییز منفی و حذف یا اصلاح آن‌ها.

جدول شاخص تمایز (Discrimination Index)

سطح تمایز	تفسیر	تعداد تقریبی سؤال	وضعیت
کمتر از ۰.۲	ضعیف	قابل توجه	نامطلوب. این سؤالات نمی‌توانند بین دانشجویان قوی و ضعیف تمایز قائل شوند.
۰.۲ تا ۰.۳	قابل قبول	چند مورد	نیاز به بهبود.
بالتر از ۰.۳	خوب	محدود	مطلوب اما کم تعداد.
منفی	بسیار نامطلوب	حداقل ۳ مورد (مانند سوال ۸۰)	بحرانی. این سؤالات به اشتباه توسط دانشجویان قوی بیشتر انتخاب می‌شوند و باید حذف گردند.

میانگین ضریب تمییز: ۰.۲۱۳ (کم)

نتیجه گیری: توانایی تمایزگذاری آزمون به طور کلی ضعیف است. وجود سؤالات با تمایز منفی (مانند سؤالات ۸۰، ۸۸ و ۹۳) به شدت کیفیت آزمون را زیر سؤال برده و اعتبار آن را خدشه‌دار می‌کند. البته برای آزمونهای رقابتی این چنین است. اما ب رای آزمونهای پیشرفت تحصیلی کافی است ضریب تمییز منفی نباشد.

۵. اعتبارسنجی KR20 و آلفای کرونباخ

- مقدار: (۰.۵۸۵-) ، غیرقابل قبول
- تفسیر: مقادیر منفی نشان‌دهنده ناهمگونی شدید در پاسخ‌ها و احتمالاً خطا در کلید سؤالات یا ساختار آزمون است. انتظار می‌رود با افزایش تعداد سؤالات، پایایی آزمون افزایش یابد؛ اما در این آزمون، هر دو شاخص پایایی KR-20 و آلفای کرونباخ مقدار منفی دارند. پایایی منفی به معنای وجود ناهمخوانی شدید بین سؤالات و نمره کل است. دلایل اصلی آن به شرح زیر است:
- ۱- آسانی بیش از حد سؤالات و کاهش واریانس کل آزمون
- ۲. ضعف در ضریب تمییز که نشان‌دهنده عدم همبستگی مثبت میان آیتم‌ها و توانایی کلی دانشجویان است
- ۳. احتمال خطا در کلید تصحیح یا اشتباه در وارد کردن پاسخ صحیح برخی سؤالات
- ۴. عدم همگونی محتوایی آزمون (تنوع موضوعات تدریس شده، تنوع سؤالاتی که مربوط به سازه‌های ذهنی ناهمگنی هستند)، به نحوی که برخی سؤالات همسو با هدف کلی آزمون نبوده‌اند.

• پیشنهاد:

- بازبینی کلید سؤالات
- بررسی ساختار آزمون از نظر همگنی سؤالات
- استفاده از تحلیل عاملی برای بررسی سازه آزمون

۶. خطای استاندارد اندازه‌گیری (SEM)

- مقدار: ۰.۰۹۳ ، پایین

- تفسیر: پایین بودن SEM به دلیل پایین بودن انحراف معیار است. اگرچه پایین بودن SEM مطلوب است، اما در اینجا نشان دهنده عدم تنوع در نمرات است.
- پیشنهاد: افزایش تنوع سؤالات به منظور افزایش دامنه نمرات.

۷. توزیع سؤالات از نظر دشواری

- آسان: ۸۲٪
- مناسب: ۱۲٪
- دشوار: ۶٪
- تفسیر: عدم تعادل آشکار در سطح دشواری سؤالات.
- پیشنهاد: هدف گذاری برای توزیع متعادل: ۳۰٪ آسان، ۴۰٪ متوسط، ۳۰٪ دشوار.

۸. گزینه های انحرافی (Distractors)

- برخی گزینه های انحرافی اصلاً انتخاب نشده اند. این امر بیانگر ضعف طراحی گزینه هاست؛ چون گزینه انحرافی مؤثر باید حداقل درصدی از داوطلبان ضعیف تر را به خود جذب کند. وجود گزینه های بی اثر باعث کاهش کارایی سؤال و افت ضریب تمییز می شود.
- تحلیل: در بسیاری از سؤالات (مانند سؤال ۲)، گزینه های انحرافی به خوبی طراحی نشده اند و درصد انتخاب بسیار پایینی دارند.
 - مثال: در سؤال ۲، گزینه A تنها ۵.۸۸٪ انتخاب شده که نشان می دهد این گزینه جذابیت لازم را ندارد.
 - پیشنهاد: بازبینی گزینه های انحرافی به منظور افزایش جذابیت و عملکرد آنها.

۹. تحلیل شاخص Omit

- تعریف: درصد شرکت کنندگانی که به یک سؤال پاسخی نداده اند.
 - یافته ها: در بعضی سؤالات، مقدار Omit بیش از حد معمول است. این موضوع نشان می دهد سؤال ممکن است:
 - مبهم یا طولانی باشد.
 - بسیار دشوار طراحی شده باشد.
 - یا داوطلبان به دلیل کمبود وقت از پاسخ گویی صرف نظر کرده باشند.
 - تفسیر: سؤالات با Omit بالا نیاز به بازبینی دارند؛ زیرا بی پاسخ ماندن زیاد یک سؤال، اعتبار آن را زیر سؤال می برد.
- معنی و تفسیر شاخص Omit :
- مقدار بالا (مثلاً بیش از ۵٪): ممکن است نشان دهنده مشکلی در سؤال باشد (مثلاً سؤال مبهم، بسیار دشوار، یا زمان بندی نامناسب آزمون).
 - مقدار پایین (نزدیک به صفر): نشان می دهد که اکثر شرکت کنندگان تلاش کرده اند به سؤال پاسخ دهند.

دلایل احتمالی برای Omit بالا:

۱. سؤال بیش از حد دشوار است.
۲. سؤال مبهم یا گمراه کننده است.
۳. زمان کافی برای پاسخ دهی وجود نداشته است.
۴. بی علاقه گی یا خستگی شرکت کنندگان.

چگونه با Omit برخورد کنیم؟

- اگر omit یک سؤال بالا باشد، باید سؤال را بازبینی کرد (کاهش سختی، یا حذف آن).
- اگر omit پایین است، معمولاً جای نگرانی نیست.

مثال:

در گزارش آزمون، برای هر سؤال یک ستون به نام Omit وجود دارد که تعداد افرادی را که پاسخ نداده‌اند، نشان می‌دهد (به صورت عددی یا درصدی). به عنوان مثال:

- سؤال ۱: $Omit = 1$ یک نفر پاسخ نداده است.
- سؤال ۵: $Omit = 0$ همه پاسخ داده‌اند.
- سؤال ۹۷: $Omit = 1$ یک نفر پاسخ نداده است.

در این آزمون، مقادیر Omit عموماً پایین هستند (۰ یا ۱ نفر)، که نشان می‌دهد اکثر شرکت‌کنندگان به سؤالات پاسخ داده‌اند و احتمالاً مشکل حادی از نظر درک سؤالات یا زمان‌بندی وجود نداشته است.

آیا Omit در پایایی آزمون تأثیر دارد؟

بله، اگر تعداد زیادی از شرکت‌کنندگان به یک سؤال پاسخ ندهند، می‌تواند بر شاخص‌های روانسنجی (مانند ضریب سختی، تمییز، و پایایی) تأثیر بگذارد. اما در مورد آزمون شما، از آنجا که مقادیر Omit کم هستند، به نظر نمی‌رسد نقش عمده‌ای در پایایی منفی آزمون داشته باشند.

۱۰. تحلیل شاخص RPbis

- تعریف: ضریب همبستگی نقطه‌ای دو وجهی که رابطه پاسخ درست/غلط به یک سؤال را با نمره کل آزمون می‌سنجد.
- یافته‌ها: بسیاری از سؤالات دارای RPbis بسیار پایین یا حتی منفی بوده‌اند.
- تفسیر:
 - RPbis مثبت و بالا (≥ 0.3): نشان‌دهنده کیفیت خوب سؤال.
 - RPbis نزدیک صفر: سؤال بی‌اثر از نظر تمییز.
 - RPbis منفی: نشان‌دهنده خطای کلید یا ضعف طراحی (چون دانشجویان ضعیف بیشتر از قوی‌ها درست زده‌اند).
- نتیجه: مقادیر پایین و منفی RPbis در این آزمون بیانگر ضعف کلی سؤالات در سنجش واقعی توانایی دانشجویان است.

مثال:

در گزارش آزمون، برای هر سؤال یک مقدار RPbis ارائه شده است. به عنوان مثال:

- سؤال ۱: $RPbis = 0.44 \rightarrow$ مناسب
- سؤال ۲: $RPbis = 0.06 \rightarrow$ کم (ضعیف)
- سؤال ۳: $RPbis = 0.94 \rightarrow$ عالی

چرا RPbis منفی می‌شود؟

اگر مقدار RPbis سؤالی منفی باشد، نشان دهنده آن است که افرادی که نمره کل پایین‌تری دارند، بیشتر به آن سؤال پاسخ صحیح داده‌اند. این امر می‌تواند به دلایل زیر باشد:

- سؤال مبهم یا گمراه‌کننده است.
- سؤال سطح دشواری نامناسبی دارد.
- ممکن است خطای تصادفی در پاسخ‌دهی رخ داده باشد.

۱۱. تحلیل گزینه‌های انحرافی که توسط هیچ کدام از شرکت کنندگان انتخاب نشده‌اند:

انتخاب نشدن برخی گزینه‌های انحرافی توسط هیچ یک از شرکت کنندگان نشان‌دهنده ضعف در طراحی گزینه‌های انحرافی آن سؤالات است. این مسئله می‌تواند به دلایل زیر رخ داده باشد:

دلایل احتمالی:

۱. گزینه‌های انحرافی غیرواقعی یا بسیار دور از ذهن هستند؛ به گونه‌ای که حتی شرکت کنندگان ضعیف نیز آنها را انتخاب نمی‌کنند.
۲. گزینه صحیح به قدری آشکار است که دیگر گزینه‌ها جذابیتی ندارند.
۳. خطای در طراحی سؤال (مثلاً گزینه‌های انحرافی مرتبط با موضوع نیستند).

تأثیر بر آزمون:

- کاهش قدرت تمییز سؤال: اگر گزینه‌های انحرافی جذاب نباشند، سؤال نمی‌تواند به خوبی بین افراد قوی و ضعیف تمایز قائل شود.
- پایایی و اعتبار آزمون را کاهش می‌دهد: سؤالات با گزینه‌های انحرافی ضعیف، معمولاً همبستگی پایینی با نمره کل دارند (RPbis پایین).
- سؤال بیش از حد آسان می‌شود: اگر فقط یک گزینه منطقی وجود داشته باشد، حتی شرکت کنندگان ضعیف نیز به راحتی پاسخ صحیح را حدس می‌زنند.

راهکارهای بهبود:

۱. بازنگری گزینه‌های انحرافی: گزینه‌ها باید جذاب و نزدیک به پاسخ صحیح باشند تا افراد بی‌دقت یا دارای دانش ناقص را به دام بیندازند.
 ۲. استفاده از اشتباهات رایج به عنوان گزینه انحرافی: مثلاً اگر دانشجویان معمولاً در محاسبه نوع خاصی از فرمول اشتباه می‌کنند، از نتیجه آن اشتباه به عنوان گزینه انحرافی استفاده شود.
 ۳. حذف یا جایگزینی گزینه‌های بی‌استفاده: اگر گزینه‌ای توسط هیچ‌کس انتخاب نمی‌شود، آن را با گزینه‌ای واقع‌بینانه‌تر جایگزین کنید.
- مثال: در سؤالاتی مانند سؤال ۵ (گزینه‌های A, B, D انتخاب نشده‌اند) یا سؤال ۱۰ (هیچ کدام از گزینه‌های انحرافی انتخاب نشده‌اند)، این مشکل به وضوح دیده می‌شود. این سؤالات باید به طور جدی بازبینی شوند.
- جمع‌بندی: گزینه‌های انحرافی انتخاب نشده نشان‌دهنده ضعف در طراحی آزمون است و باید اصلاح شوند تا آزمون از نظر روانسنجی معتبرتر شود.

۱۲. نتیجه‌گیری کلی

آزمون حاضر از نظر محتوایی ممکن است غنی باشد، اما از نظر روانسنجی دارای ضعف‌های جدی است، به‌ویژه در زمینه اعتبار و تمایزگذاری. پایین بودن انحراف معیار و واریانس نیز نشان‌دهنده عدم توانایی آزمون در تفکیک دانشجویان است.

۱۳. پیشنهادات کلی

- بازبینی کامل کلید سؤالات و ساختار آزمون
- کاهش سؤالات بسیار آسان و افزایش سؤالات با دشواری متوسط
- اصلاح سؤالات با ضریب تمییز منفی
- استفاده از نظرات متخصصان برای بهبود گزینه‌های انحرافی
- بازنگری در طراحی سؤالات با توجه به شاخص‌های Omit و RPbis.
- تقویت گزینه‌های انحرافی به گونه‌ای که برای داوطلبان ضعیف‌تر، جذابیت بیشتری داشته باشند.
- حذف یا بازنویسی سؤالات با RPbis منفی یا ضریب تمییز پایین.
- تنظیم توزیع دشواری سؤالات (۳۰٪ آسان، ۴۰٪ متوسط، ۳۰٪ دشوار).

علت‌های احتمالی منفی شدن ضرایب آلفای کرانباخ و اعتبارسنجی KR20 چیست؟

منفی شدن ضرایب آلفای کرانباخ و KR-20 یک اتفاق نادر و حیاتی است که نشان‌دهنده وجود یک مشکل ساختاری جدی در داده‌های آزمون است. این وضعیت هرگز در یک آزمون سالم رخ نمی‌دهد. در مورد این آزمون، با مقادیر منفی ۹.۵۸۵، ما با یک مشکل اساسی روبرو هستیم. در ادامه علل احتمالی این پدیده را به ترتیب احتمال وقوع توضیح می‌دهیم:

۱. خطا در کلید پاسخ (Answer Key Error) محتمل‌ترین علت

این رایج‌ترین دلیل برای منفی شدن آلفا است.

- **توضیح:** وقتی شما به نرم‌افزار تحلیل می‌گویید که گزینه ۱ پاسخ صحیح است، اما در واقعیت گزینه ۳ صحیح است، نرم‌افزار تمام افرادی که گزینه ۳ را انتخاب کرده‌اند، "غلط" و تمام افرادی که گزینه ۱ زده‌اند را "صحیح" محاسبه می‌کند.

- **تأثیر:** در این حالت، دانشجویان قوی که پاسخ درست (گزینه ۳) داده‌اند، نمره نمی‌گیرند و دانشجویان ضعیف که به طور تصادفی گزینه ۱ را انتخاب کرده‌اند، نمره می‌گیرند. این باعث می‌شود همبستگی بین سؤالات و نمره کل معکوس شود (یعنی هرچه فرد قوی‌تر باشد، نمره کمتری می‌گیرد). همین همبستگی معکوس است که باعث می‌شود واریانس کل منفی شده و آلفا منفی شود.

۲. معکوس بودن سیستم نمره‌دهی (Coding Error)

- **توضیح:** ممکن است در هنگام وارد کردن داده‌ها، مقادیر "صحیح" و "غلط" معکوس کدگذاری شده باشند. برای مثال، به جای این که عدد ۱ برای "صحیح" و ۰ برای "غلط" در نظر گرفته شود، عدد ۰ برای "صحیح" و ۱ برای "غلط" ثبت شده باشد.

- **تأثیر:** دقیقاً مانند مورد اول، این کار باعث می‌شود عملکرد دانشجویان قوی و ضعیف معکوس تفسیر شود و منجر به محاسبه آلفای منفی گردد.

۳. وجود سؤالات بسیار معیوب با تمییز منفی (Items with Strong Negative Discrimination)

- **توضیح:** اگر چندین سؤال (نه فقط یکی دو تا) به شدت معیوب باشند به طوری که ضریب تمییز (D) یا همبستگی نقطه‌ای (RPbis) آنها منفی بسیار قوی باشد، این می‌تواند کل اعتبار آزمون را به زیر صفر بکشد.

- **چگونه؟** سؤالی که تمییز منفی قوی دارد، نه تنها کمکی به اعتبار نمی‌کند، بلکه به طور سیستماتیک آن را کاهش می‌دهد. اگر تعداد این سؤالات از یک حدی بیشتر شود، می‌توانند منجر به آلفای منفی شوند. در گزارش شما سؤالات ۸۰، ۸۸ و ۹۳ تمییز منفی دارند که می‌توانند مقصر باشند، اما معمولاً این تعداد سؤال برای ایجاد آلفای منفی به این شدت کافی نیست و حتماً باید علت اصلی‌تری مانند مورد اول وجود داشته باشد.

۴. واریانس بسیار پایین نمرات (Extremely Low Variance)

- **توضیح:** آلفای کرانباخ و KR-20 به واریانس نمرات وابسته هستند. **اگر واریانس نمرات به صفر نزدیک شود (یعنی همه نمرات تقریباً یکسان باشد)**، مخرج کسر در فرمول محاسبه آلفا به نزدیک صفر می‌رسد و در نتیجه مقدار آلفا می‌تواند به صورت تئوری به بی‌نهایت (و منفی) میل کند.

- **تأثیر:** در این آزمون، انحراف معیار ۰.۹۲۹ است که پایین است، اما به تنهایی برای ایجاد آلفای منفی ۹.۵۸۵ کافی نیست. این عامل معمولاً در کنار یکی از عوامل بالا (به ویژه مورد ۱ و ۲) تشدید کننده است.

۵. ناهمگونی مفرط سؤالات (Extreme Heterogeneity)

- **توضیح:** آلفا هماهنگی درونی بین سؤالات را اندازه می‌گیرد. اگر سؤالات آنقدر ناهمگون باشند که در واقع چندین سازه کاملاً متفاوت (مثلاً همزمان سؤالات ریاضی، ادبیات و تاریخ) را اندازه بگیرند، می‌توانند منجر به همبستگی منفی بین سؤالات شوند.

- **تأثیر:** این مورد کمتر شایع است و معمولاً به تنهایی باعث منفی شدن آلفا نمی‌شود، مگر اینکه با سایر خطاها ترکیب شود.

خلاصه و راهکار فوری برای تشخیص علت:

بهترین و سریع ترین راه برای تشخیص علت، انجام این دو کار است:

۱. بررسی فوری کلید پاسخ: همه سؤالات را یکی یکی با کلید پاسخ تطبیق دهید. این محتمل ترین علت است.

۲. خروجی گرفتن ماتریس همبستگی بین سؤالات: از نرم افزار بخواهید ماتریس همبستگی بین تک تک سؤالات را چاپ کند. اگر تعداد زیادی از همبستگی ها منفی هستند (به ویژه همبستگی های قوی منفی)، مطمئن ترین نشانه وجود خطای کلید پاسخ یا کدگذاری معکوس است.

جمع بندی نهایی: با توجه به شدت منفی بودن مقادیر (۰.۵۸۵-)، تقریباً می توان با اطمینان گفت که علت اصلی یا خطا در کلید پاسخ است یا یک اشتباه فاحش در کدگذاری داده ها. تا زمانی که این مشکل ریشه ای برطرف نشود، تحلیل سایر شاخص ها (مانند دشواری و تمیز) نیز بی اعتبار است، زیرا این شاخص ها نیز بر اساس همان کلید پاسخ غلط محاسبه شده اند.

واریانس ایده آل یک آزمون چقدر باید باشد؟

پاسخ به این سؤال نیاز به درک رابطه بین واریانس و نمره کل آزمون دارد.

پاسخ کوتاه:

واریانس ایده آل برای یک آزمون، واریانسی است که نشان دهنده گستره مناسبی از توانایی های شرکت کنندگان باشد و به آزمون اجازه دهد تا به خوبی بین افراد تمایز قائل شود. این مقدار مطلق ثابتی ندارد و به مقیاس نمره دهی آزمون بستگی دارد.

پاسخ دقیق و علمی:

برای تعیین واریانس ایده آل، باید آن را در نسبت با حداکثر واریانس ممکن برای آن آزمون سنجید. حداکثر واریانس زمانی رخ می دهد که نیمی از شرکت کنندگان کمترین نمره (صفر) و نیمی دیگر بیشترین نمره (نمره کل) را کسب کنند.

فرمول محاسبه واریانس ایده آل:

حداکثر واریانس نظری (σ^2_{max}) برای یک آزمون از رابطه زیر به دست می آید:

$$\sigma^2_{max} = (2 / \text{نمره کل})^2$$

نمره کل آزمون شما: برابر ۱۳ است، بنابراین، حداکثر واریانس نظری برای این آزمون برابر است با:

$$(13 / 2)^2 = (6.5)^2 = 42.9$$

حالا چگونه این عدد را تفسیر کنیم؟

۱. واریانس واقعی آزمون شما: ۰.۸۶۳.

۲. حداکثر واریانس ممکن: ۴۲.۹.

مشاهده می کنید که واریانس آزمون شما (۰.۸۶۳) نسبت به حداکثر ممکن (۴۲.۹) بسیار ناچیز است. این نشان می دهد که تقریباً تمام نمرات در یک نقطه متمرکز شده اند و آزمون توانایی بسیار کمی در تمایز بین دانشجویان دارد.

یک قاعده سرانگشتی و کاربردی:

از آنجایی که رسیدن به حداکثر واریانس نظری در عمل غیرممکن است، یک هدف واقع‌بینانه‌تر این است که واریانس آزمون حدود ۲۰٪ تا ۲۵٪ از حداکثر واریانس نظری باشد.

محاسبه برای این آزمون (نمره کل = ۱۳):

• حد پایین ایده‌آل: $42.9 \times 0.20 = 8.58$

• حد بالا ایده‌آل: $42.9 \times 0.25 = 10.72$

نتیجه‌گیری برای این آزمون:

• واریانس فعلی: ۰.۸۶۳

• محدوده واریانس ایده‌آل: ۸.۵۸ تا ۱۰.۷۲

واریانس این آزمون (۰.۸۶۳) به شدت پایین‌تر از محدوده مطلوب است. این عدد به طور مستقیم تأیید می‌کند که:

• انحراف معیار پایین است (۰.۹۲۹)

• آزمون بسیار آسان است (میانگین نمرات ۱۰.۷۶ از ۱۳).

• نمرات اکثر شرکت‌کنندگان به هم نزدیک و پراکندگی کمی وجود دارد.

چگونه واریانس را افزایش دهیم؟

برای رسیدن به یک واریانس ایده‌آل (۱۰-۸)، باید توزیع دشواری سؤالات را اصلاح کرد:

۱. کاهش سؤالات بسیار آسان: سؤالاتی که بیش از ۸۰٪ افراد به آنها پاسخ صحیح داده‌اند (مانند سؤالات ۳، ۴، ۵، ۱۰) را بررسی و تعدیل کنید.

۲. افزایش سؤالات با دشواری متوسط: هدف نهایی داشتن حدود ۶۰٪ سؤالات با ضریب دشواری بین ۰.۴ تا ۰.۷.

۳. افزودن سؤالات دشوارتر: حدود ۲۰٪ سؤالات باید به گونه‌ای باشند که تنها افراد بسیار قادر بتوانند به آنها پاسخ دهند (ضریب دشواری زیر ۰.۴).

با انجام این اصلاحات، نمرات از هم فاصله می‌گیرند، انحراف معیار افزایش می‌یابد و در نتیجه واریانس به محدوده مطلوب نزدیک می‌شود. این کار به طور خودکار اعتبار (آلفا) آزمون را نیز dramatically افزایش خواهد داد.

آلفای کرانباخ چقدر باشد؟

مقدار آلفای کرانباخ یک شاخص کلیدی برای سنجش کیفیت یک آزمون است، اما تفسیر آن بستگی به هدف و حساسیت آزمون دارد. در ادامه یک راهنمای جامع و کاربردی ارائه می‌شود:

مقدار آلفا بین ۰ تا ۱ است. تفسیر کلی آن به صورت زیر است:

محدوده آلفای کرانباخ	تفسیر	وضعیت برای این آزمون = ۹.۵۸-
$\alpha \geq 0.9$	عالی (Excellent)	غیر قابل قبول
$0.8 \leq \alpha < 0.9$	خوب - (Good) برای اکثر آزمون‌ها کافی است.	
$0.7 \leq \alpha < 0.8$	قابل قبول - (Acceptable) برای اهداف پژوهشی مقدماتی.	
$0.6 \leq \alpha < 0.7$	مشکوک - (Questionable) نیاز به بازبینی دارد.	
$\alpha < 0.6$	غیر قابل قبول - (Unacceptable) آزمون فاقد قابلیت اعتماد است.	
آلفا کوچکتر از صفر یا منفی	خطای فاحش - (Critical Error) نشان‌دهنده یک مشکل ساختاری جدی است.	

جزئیات بیشتر بر اساس نوع آزمون:

• آزمون‌های حساس و High-Stakes (مانند کنکور، گواهینامه‌های حرفه‌ای):

○ مقدار مطلوب: $\alpha \geq 0.9$

○ دلیل: این آزمون‌ها قرار است سرنوشت‌ساز باشند، بنابراین باید از حداکثر پایداری و ثبات برخوردار باشند.

• آزمون‌های استاندارد پیشرفت تحصیلی (مثل امتحانات پایانی دانشگاه):

○ مقدار مطلوب: $\alpha \geq 0.8$

○ دلیل: این آزمون‌ها برای نمره‌دهی و تصمیم‌گیری آموزشی استفاده می‌شوند و نیاز به قابلیت اعتماد بالا دارند.

• پرسشنامه‌های تحقیقاتی و پژوهشی (مثلاً در مطالعات روانشناسی):

○ مقدار قابل قبول: $\alpha \geq 0.7$

○ توجه: برای مقیاس‌های جدید یا اکتشافی، حتی مقادیر بین ۰.۶ تا ۰.۷ نیز ممکن است تحمل شود، اما ایده‌آل نیست.

نکات حیاتی در تفسیر آلفا:

۱. آلفای منفی (مورد شما): همانطور که در جدول بالا دیدید، آلفای منفی هرگز طبیعی نیست. این وضعیت نشان‌دهنده یکی از مشکلات زیر است:

*خطا در کلید پاسخ (محتمل‌ترین علت)

*معکوس شدن داده‌ها (مثلاً صفر و یک جابجا شده باشند)

*وجود سؤالات بسیار معیوب با تمییز منفی قوی که کل محاسبات را بر هم زده است.

۲. آلفای خیلی بالا (مثلاً بالای ۰.۹۵): اگرچه به نظر خوب می‌رسد، اما می‌تواند نشان‌دهنده تکرار غیرضروری سؤالات (redundancy) باشد. یعنی سؤالات آنقدر شبیه به هم هستند که اطلاعات جدیدی اضافه نمی‌کنند. گاهی اوقات حذف ۱-۲ سؤال می‌تواند آلفا را کاهش دهد اما آزمون را کارآمدتر کند.

۳. تعداد سؤالات: آلفا به تعداد سؤالات آزمون حساس است. آزمون‌های کوتاه‌تر (مثلاً زیر ۱۰ سؤال) به طور طبیعی آلفای پایین‌تری دارند. آزمون شما با صد سؤال، انتظار می‌رود آلفای بسیار بالایی داشته باشد.

جمع‌بندی برای این آزمون:

مقدار $\alpha = -9.585$ به وضوح نشان می‌دهد که داده‌های این آزمون برای محاسبه اعتبار غیرقابل استفاده هستند. قبل از هرگونه تحلیل دیگری، باید منبع این خطا را پیدا و رفع کنید. اولین و مهم‌ترین گام، بازبینی کامل کلید پاسخ و مطمئن شدن از صحت کدگذاری داده‌هاست. تا زمانی که این مشکل رفع نشود، سایر شاخص‌های گزارش شده (مانند دشواری و تمییز) نیز قابل اعتماد نیستند.

تهیه و تنظیم: دکتر علی شیخیان