

به نام خدا

راهنمای کامل
آزمون در زبان انگلیسی
نظریه و کاربردها

کد کتاب ۱۱۱

نویسندگان
دکتر حسین فرهادی
دکتر عبدالجواد جعفرپور
دکتر پرویز بیرجندی

مترجمین: راحله قنبری

علی نقدی

عنوان : راهنمای کامل آزمون زبان انگلیسی

مترجمین : رامله قنبری - علی نقدی

انتشارات : مهکامه

چاپ و صمافی : میلاد نور

نوبت چاپ : دوم / ۱۳۸۷

تیراژ : ۲۰۰۰

صمفه آرایى : گروه هنرى (ز)

شماره شابک : 964-96325-2-2

قیمت : ۷۰۰۰۰ (ریال)

قنبرى . راحله . ۱۳۵۱-

راهنمای کامل آزمون در زبان انگلیسی : نظریه ها و کاربردها /
مترجمین راحله قنبرى . علی نقدی . تهران : مهکامه . ۱۳۸۴ .
۳۳۲ ص .

ISBN:964-96325-2-2

فهرست نویسى براساس اطلاعات فیپا .

کتاب حاضر راهنمای کتاب "Testing language skill

from theory to practice" تألیف حسین فرهادی . عبدالجواد

جعفرپور و پرویز بیرجندی است .

۱ . زبان -- آزمونها . ۲ . زبان انگلیسی -- آزمونها . ۳ . زبان

انگلیسی -- حسین . ۱۳۲۵ - ج . جعفرپور . عبدالجواد . ۱۳۲۶ - د .

بیرجندی . پرویز . ۱۳۱۹ - ه . عنوان .

۴۵۲۶ ت ۴ / ۶ / ۵۳ P ۴۱۸ / ۰۰۷۶

۲۶۳۱۵ - ۸۴ م

کتابخانه ملی ایران

هرگونه کپی برداری و استفاده از مطالب این کتاب بدون مجوز کتبی ناشر ممنوع بوده و پیگرد
قانونی خواهد داشت .

مرکز پخش :

تهران - فیابان انقلاب - فیابان ۱۲ فروردین - کوچه بهشت آیین - پلاک ۳۱ - طبقه سوم

تلفن : ۶۶۹۶۱۵۰۹ - ۶۶۴۹۷۰۵۰

فهرست مطالب

فصل اول

۱۱.....	مقدمات
۱۱.....	۱-۱ اهمیت آزمون
۱۲.....	۱-۲ سیاست‌گذاری و تصمیم‌گیری
۱۳.....	۱-۳ تست، سنجش و ارزیابی
۱۳.....	۱-۴ آزمون زبان
۱۴.....	۱-۵ آزمون برای چیست؟
۱۵.....	۱-۶ آزمون‌های معلم ساخته در مقایسه با آزمون‌های استاندارد
۱۷.....	۱-۷ آموزش و آزمون زبان
۱۷.....	۱-۷-۱ روش آزمون‌های قدیمی
۱۹.....	۱-۷-۲ تست‌های چند گزینه‌ای
۲۱.....	۱-۷-۳ آزمون ارتباطات (زبان مکالمه)
۲۲.....	فعالیت‌ها

فصل دوم

۲۳.....	نقش آزمون زبان
۲۴.....	۲-۱ مقدمه
۲۵.....	۲-۲ تست‌های پیش‌بین
۲۵.....	۲-۲-۱ آزمون‌هایی برای گزینش
۲۷.....	۲-۲-۲ آزمون‌های رتبه‌بندی (تعیین سطح)
۲۸.....	۲-۲-۳ تست‌های تشخیص استعداد
۲۹.....	۲-۳ نمره‌گذاری تست‌های تحقیقی
۲۹.....	۲-۳-۱ تست‌های تحقیقی
۳۰.....	۲-۳-۲ تست‌های تخصصی
۳۲.....	۲-۳-۳ تست‌های علوم

فصل سوم

۳۳.....	اشکال آزمون زبان
۳۳.....	۳-۱ مقدمه

۳۴	۳-۲ ساختار یک آیتم
۳۵	۳-۳ طبقه‌بندی اشکال آیتم
۳۵	۳-۳-۱ آیتم‌های ذهنی و عینی
۳۷	۳-۳-۲ نارسائی تست تشریحی در مقابل آیتم‌های چندگزینه‌ای
۳۸	۳-۳-۳ آیتم‌های کامل کردنی در مقابل آیتم‌های تشخیصی
۳۸	۳-۴ طبقه‌بندی زبان - روانی
۴۱	فعالیت‌ها:

فصل چهارم

۴۲	روش‌های اساسی آمار در آزمون
۴۲	۴-۱ مقدمه
۴۳	۴-۲ اقدام به عمل برحسب توصیه‌های آماری
۴۴	۴-۳ جدول‌بندی داده‌ها
۴۴	۴-۳-۱ به‌ترتیب رتبه
۴۵	۴-۳-۲ توزیع فراوانی
۴۷	۴-۳-۳ تشخیص درصدی
۴۷	۴-۳-۳-۱ فراوانی نسبی
۴۸	۴-۳-۳-۲ فراوانی کلی
۵۰	۴-۴ تنظیم گرافیکی داده‌ها
۵۱	۴-۴-۱ ترسیم توزیع فراوانی
۵۱	۴-۴-۱-۱ نمودار ستونی
۵۲	۴-۴-۱-۲ توزیع بطریق چهار وجهی
۵۳	۴-۴-۱-۳ چند وجهی‌های فراوانی
۵۳	۴-۵ آمارهای توضیحی
۵۴	۴-۵-۱ شاخص‌های گرایش محوری
۵۴	۴-۵-۱-۱ مد نمرات
۵۴	۴-۵-۱-۲ میانگین
۵۶	۴-۵-۲ شاخص‌های مربوط به متغیرها (تغییرات)
۵۶	۴-۵-۲-۱ دامنه
۵۷	۴-۵-۲-۲ SD و واریانس
۵۹	۴-۵-۲-۳ محاسبه واریانس با استفاده از نمرات خام
۶۱	۴-۶ همبستگی
۶۲	۴-۶-۱ بسنگی خطی
۶۵	۴-۶-۲ ضریب همبستگی
۶۹	۴-۶-۳ بسنگی به ترتیب رتبه
۷۰	۴-۶-۴ همبستگی

فصل پنجم

۷۵	ساختار آزمون
۷۵	۵-۱ مقدمه
۷۶	۵-۲ چگونگی شکل و کار یک آزمون

۵-۳ طرح و برنامه‌ریزی (تعیین موضوع آزمون).....	۷۷
۵-۴ تهیه و تنظیم آیتما.....	۷۹
۵-۴-۱ پیشنهاداتی پیرامون سئوالات درست یا غلط.....	۷۹
۵-۴-۲ پیشنهاداتی مبنی بر تهیه آیتماهای جور کردنی.....	۸۱
۵-۴-۳ نظراتی پیرامون تهیه آیتماهای چند گزینه‌ای.....	۸۲
۵-۵ تجدیدنظر (بازبینی).....	۸۶
۵-۶-۱ آسان بودن یک آیتم.....	۸۷
۵-۶-۲ خاصیت فرق نمایی آیتم (افتراق سوال).....	۸۹
۵-۶-۳ توزیع گزینه‌ها.....	۹۱
۵-۷ اصول عقل و منطق را رعایت کنید.....	۹۲

فصل ششم

مشخصات یک آزمون خوب.....	۹۳
۶-۱ مقدمه.....	۹۴
۶-۲ پایایی.....	۹۵
۶-۳ شیوه‌های برآورد ضریب پایایی یک آزمون.....	۱۰۱
۶-۳-۱ شیوه آزمون و آزمون مجدد.....	۱۰۲
۶-۳-۲ شیوه اشکال همسو.....	۱۰۳
۶-۳-۳ شیوه دو نیمی Split-half.....	۱۰۳
۶-۳-۴ شیوه KR-21.....	۱۰۵
۶-۴ عوامل مؤثر بر پایایی.....	۱۰۶
۶-۴-۱ تأثیرگذاری آزمون‌دهنده.....	۱۰۷
۶-۴-۲ تأثیرات عوامل آزمون.....	۱۰۸
۶-۴-۲-۱ ساخت آزمون.....	۱۰۸
۶-۴-۲-۲ اثر عوامل اجرایی.....	۱۱۲
۶-۴-۲-۳ اثر عوامل نمره‌گذار.....	۱۱۲
۶-۵ اعتبار بخشیدن (اعتبار).....	۱۱۳
۶-۵-۱ اعتبار موضوعی.....	۱۱۴
۶-۵-۲ معتبر بودن بر پایه معیار.....	۱۱۶
۶-۵-۲-۱ اعتبار حاصل از همزمانی امتحانی.....	۱۱۶
۶-۵-۲-۲ پیش‌بینی معتبر.....	۱۱۶
۶-۵-۳ معتبر بودن ساختاری.....	۱۱۷
۶-۶ عواملی که بر معتبر بودن آزمون اثر می‌گذارند.....	۱۱۸
۶-۶-۱ دستورالعمل راهنما.....	۱۱۸
۶-۶-۲ مشکل بودن سطح آزمون.....	۱۱۸
۶-۶-۳ ساخت آیتما.....	۱۱۹
۶-۶-۴ تنظیم آیتما و پاسخ‌های درست.....	۱۱۹
۶-۷ پیوند بین قابل اعتبار بودن و قابل انکاء بودن.....	۱۱۹
۶-۸ پایایی، اعتبار و مقبول بودن.....	۱۲۰
۶-۹ قابل اجرا بودن در عمل.....	۱۲۱
۶-۹-۱ سادگی در اجرا.....	۱۲۱

- ۱۲۲ حفظ سادگی در نمره دادن ۶-۹-۲
 ۱۲۲ سهولت تفسیر و کاربرد ۶-۹-۳

فصل هفتم

- آزمون زبان و نظریه‌هایی درباره آن ۱۲۵
 مقدمه ۱۲۵
 ۷-۲ روش نکات تجریدی (مجزا شکل) ۱۲۷
 ۷-۳ روش جامع ۱۲۸
 ۷-۴ روش عملی ۱۳۰
 ۷-۵ تفسیر نمرات آزمون ۱۳۲
 ۷-۵-۱ تفسیر متعارف ۱۳۳
 ۷-۵-۲ تفسیر ملاکی ۱۳۴

فصل هشتم

- آزمون واژگان ۱۳۷
 ۸-۱ مقدمه ۱۳۷
 ۸-۲ رهنمودی برای تألیف سئوالات واژگان ۱۳۹
 فعالیت‌ها ۱۴۱

فصل نهم

- آزمون گرامری ۱۴۴
 ۹-۱ مقدمه ۱۴۴
 ۹-۲ رهنمودی برای تهیه و تنظیم سئوالات گرامر ۱۴۶

فصل دهم

- آزمون طرز تلفظ ۱۵۰
 ۱۰-۱ مقدمه ۱۵۰
 ۱۰-۲ تشخیص ۱۵۱
 ۱۰-۲-۱ تمایز صوتی ۱۵۱
 ۱۰-۲-۲ تشخیص تکیه ۱۵۲
 ۱۰-۲-۳ تشخیص لحن صدا ۱۵۳
 ۱۰-۳ ادای واج‌ها ۱۵۴
 ۱۰-۳-۱ تقلید ۱۵۴
 ۱۰-۳-۲ با صدای بلند خواندن ۱۵۵
 ۱۰-۳-۳ بازگوئی (واگوئی) ۱۵۵
 ۱۰-۳-۴ صحبت کردن درباره تصاویر ۱۵۵
 ۱۰-۴ رهنمودی برای تنظیم سئوالات تلفظی ۱۵۶

فصل یازدهم

- آزمون درک مطلب شنیداری ۱۶۰
 ۱۱-۱ مقدمه ۱۶۰

۱۱-۲	انواع سئوالات	۱۶۱
۱۱-۲-۱	عکس‌ها (تصاویر)	۱۶۲
۱۱-۲-۴	دیکته	۱۶۴
۱۱-۲-۵	دیالوگ	۱۶۴
۱۱-۲-۶	سخنرانی و گفتگو	۱۶۵
۱۱-۲-۷	پیامهائی با سروصداهاى مزاحم	۱۶۶
۱۱-۳	رهنمودی برای تألیف و تنظیم سئوالات درک مطلب شنیداری	۱۶۷
	فعالیت‌ها	۱۶۸

فصل دوازدهم

	آزمون بیان شفاهی	۱۷۰
۱۲-۱	مقدمه	۱۷۰
۱۲-۲	سنجش‌های غیرمستقیم	۱۷۰
۱۲-۲-۱	استفاده از تصاویر	۱۷۱
۱۲-۲-۲	تغییر شکل دادن	۱۷۴
۱۲-۲-۳	تبعیت از فرامین	۱۷۴
۱۲-۲-۴	بازگویی (واگویی)	۱۷۴
۱۲-۲-۵	شرح دادن	۱۷۴
۱۲-۲-۶	صحبت‌های کوتاه	۱۷۵
۱۲-۲-۷	نقش بازی کردن	۱۷۵
۱۲-۳	روش‌های مستقیم	۱۷۵
۱۲-۴	رهنمودی برای اجرای مصاحبه و ارزیابی آن	۱۷۸
	فعالیت‌ها	۱۸۰

فصل سیزدهم

	آزمون خواندن و درک مفاهیم	۱۸۲
۱۳-۱	مقدمه	۱۸۲
۱۳-۲	رهنمودهایی برای ساختن پرسشهای خواندن و درک مفاهیم	۱۸۶

فصل چهاردهم

	آزمون‌های مربوط به توانایی نوشتن	۱۹۱
۱۴-۱	مقدمه	۱۹۱
۱۴-۲	آمادگی برای نوشتن	۱۹۱
۱۴-۲-۱	رونویسی	۱۹۲
۱۴-۲-۲	هجی کردن	۱۹۲
۱۴-۲-۳	نقطه‌گذاری	۱۹۲
۱۴-۲-۴	ترکیب جملات	۱۹۳
۱۴-۲-۴	تقلیل جملات	۱۹۳
۱۴-۲-۶	تکمیل جملات	۱۹۳
۱۴-۲-۷	تغییر جملات	۱۹۴
۱۴-۲-۸	بسط جملات	۱۹۴

۱۹۴	تشخیص خطاها ۱۴-۲-۹
۱۹۵	نوشتن کنترل شده ۱۴-۳
۱۹۵	تغییر شکل جملات ۱۴-۳-۱
۱۹۵	جمله سازی به وسیله کلمات ۱۴-۳-۲
۱۹۵	جمله سازی از طریق پاسخ و پرسش ۱۴-۳-۳
۱۹۶	ساختن پاراگراف بوسیله یادداشتها ۱۴-۳-۴
۱۹۶	تشکیل پاراگراف ۱۴-۳-۵
۱۹۶	تکمیل پاراگراف ۱۴-۳-۶
۱۹۷	نوشته آزاد ۱۴-۴
۲۰۰	دستورالعملهایی برای آماده کردن آزمونهای نوشتاری ۱۴-۵
۲۰۲	فعالیتها

فصل پانزدهم

۲۰۳	آزمونهای جای خالی و املاء
۲۰۳	مقدمه ۱۵-۱
۲۰۵	فرآیند جا خالی ۱۵-۲
۲۱۰	گونه های فرآیند جا خالی ۱۵-۳
۲۱۲	نمره دهی آزمون جا خالی ۱۵-۴
۲۱۲	روش عین کلمات ۱۵-۴-۱
۲۱۳	روش کلمات قابل قبول ۱۵-۴-۲
۲۱۳	روش پاسخ توزینی ۱۵-۴-۳
۲۱۴	آزمونهای املاء ۱۵-۵
۲۱۵	املاي استاندارد ۱۵-۵-۱
۲۱۶	املاي ناقص ۱۵-۵-۲
۲۱۶	تقلید فراخوانی ۱۵-۵-۳
۲۱۷	املاء - انشاء ۱۵-۵-۴
۲۱۷	املا با سروصدا ۱۵-۵-۵
۲۱۸	نمره دهی آزمونهای املاء ۱۵-۶

فصل شانزدهم

۲۲۱	آزمونهای کارکردی
۲۲۱	مقدمه ۱۶-۱
۲۲۲	مشکلات آزمونهای مجزا شکل ۱۶-۲
۲۲۳	مشکلات آزمونهای تلفیقی ۱۶-۳
۲۲۴	مشکلات مربوط به پایایی ۱۶-۳-۱
۲۲۵	فرآیندهای طرح آزمونهای کارکردی ۱۶-۴
۲۲۶	انتخاب کارکردها ۱۶-۴-۱
۲۲۷	عوامل اجتماعی ۱۶-۴-۲
۲۲۷	موقعیت اجتماعی ۱۶-۴-۲-۱
۲۲۷	رابطه اجتماعی ۱۶-۴-۲-۲
۲۲۸	پایگاه اجتماعی ۱۶-۴-۲-۳

۲۲۸.....	۱۶-۴-۳ ملاک‌های عملکرد.....
۲۲۸.....	۱۶-۴-۴ طرح تنه سؤال.....
۲۲۹.....	پرسش.....
۲۲۹.....	ویژگیهای پرسش.....
۲۲۹.....	۱۶-۵ پیش‌آزمون.....
۲۲۹.....	۱-۱-۶۱ مرحله ۱: پیش‌آزمون گویشوران بومی زبان.....
۲۳۰.....	۲-۲-۱۶ مرحله ۲: پیش‌آزمون گویشوران غیربومی زبان.....
۲۳۰.....	۳-۲-۱۶ انتخاب گزینه‌ها.....
۲۳۱.....	۴-۲-۱۶ مرحله ۳: پیش‌آزمون فراگیران بومی و غیربومی زبان.....
۲۳۱.....	۱۶-۶ سیستم نمره‌دهی.....
۲۳۱.....	۱۶-۷ نتیجه‌گیری:.....
۲۳۲.....	فعالیت‌ها.....

بنام ایزد منان

همچون تو گلی شکفته پر ناز نیافت
در هیچ ورق شرح رخت باز نیافت

بلبل که ز عشق یکک هماواز نیافت
گل گرچه به حسن صد ورق داشت ولی

تقدم به روح بلند

مرحوم علی دهقانی
مرحوم سید علی سجادیان

کتابی که در پیش رو دارید از سلسله کتب راهنمای تخصصی دانشجویان است که انتشارات مهکامه به قلم این مترجمین به چاپ می‌رساند.

این اثر راهنمای کامل زبان انگلیسی برای دانشجویان رشته مترجمی و دبیری زبان انگلیسی در مقطع کارشناسی بعنوان منبع اصلی درس آزمون‌سازی می‌باشد. امید است ترجمه این اثر در ارائه مفاهیم بنیادی زبان انگلیسی مفید و مؤثر واقع شود.

برخی آثار که به قلم این مترجم برگردان و گردآوری شده است:

۱- راهنمای زبان انگلیسی برای دانشجویان هنرهای تجسمی (۱) و (۲)

۲- راهنمای زبان انگلیسی برای دانشجویان رشته کامپیوتر

۳- توانایی شناخت مفاهیم کامپیوتر (IT)

۴- ویندوز (Windows)

۵- ورد (Word)

۶- اکسل (Excel)

۷- اکسس (Access)

۸- پاورپوینت (PowerPoint)

۹- اینترنت (Internet)

۱۰- نمونه تست‌های ICDL

فصل اول

مقدمات

۱-۱ اهمیت آزمون

سنجش و ارزیابی از دیر باز کار ما انسان‌ها بوده است. انسان اصلاً با دغدغه‌ای همچون اندازه گرفتن و اندازه شدن، امتحان گرفتن و امتحان شدن بزرگ می‌شود. مربیان بویژه با این نوع سنجش و ارزیابی در حوزه پیشرفت درسی دانش‌آموزان سروکار دارند. بهمان نسبتی که اهداف تربیتی پیچیده شده است و تعداد دانش‌آموزان بشدت افزایش یافته‌اند، سنجیدن و امتحان کردن هم بدین لحاظ امری مهم و مشکل شده است. بعلاوه، مربیان همواره درصدد تجدیدنظر در برنامه‌های فعلی و ایجاد تحولاتی نو در آن شده‌اند.

همانطور که بسیاری از مردم بدرستی فکر میکنند که بخش عمده‌ای از امر اصلاح جامعه بر دوش مربیان نهاده شده است، بگمان ظاهر در برابر هر مسئله اجتماعی امروز کسی هست که قویاً معتقد باشد که ریشه حل مشکلات را باید در عرصه آموزش و پرورش جستجو کرد.

تعلیم و تربیت از مهمترین نهادهای اجتماعی محسوب میگردد. در کلیه کشورها به غیر از وزارت دفاع و پشتیبانی، این نهاد از جمله مؤسساتی است که بیشترین هزینه‌ها صرف آن میگردد. در پاره‌ای موارد هر شهروندی خود با مسائل آموزش و پرورش درگیر است. بیشتر از یک چهارم جمعیت ملل دنیا به مدرسه می‌روند.

سیستم آموزشی به واقع یک سیستم غول‌آسا و مسئولیت‌های آن بهمین اندازه بزرگ است و بنابراین ارزیابی فرایند آن و دست پرورده‌های این فرایند از اهمیت والائی برخوردار می‌باشد. در حقیقت

امر سنجش و ارزیابی در هر سیستم آموزشی حائز اهمیتی اساسی است. معلمان همواره تمایل دارند تا بدانند که میزان یادگیری و اندوخته دانش‌آموزانشان چقدر است. در ضمن دولت و بخش‌های خصوصی که دستمزد معلمان را می‌پردازند و دانش‌آموختگان را بعدها به‌استخدام خود درمی‌آورند سخت علاقمندند تا از توانمندیهای آنان اطلاعات دقیقی کسب کنند و سرانجام اینکه معلمان، دانش‌آموزان و مدیران و اولیاء در راستای دستیابی به اهداف تعلیم و تربیت همه فعالند. و طبیعی است که بخواهند تا از میزان دست یافتن به این اهداف باخبر گردند. امتحان کردن و ارزیابی ابزاری ضروری جهت کمک به رسیدن به اکثر این اهداف و بمنظور اتخاذ تصمیمات آموزشی دقیق می‌باشند.

۲-۱ سیاست‌گذاری و تصمیم‌گیری

ارتباط مستقیم هر یک از افراد در امر تعلیم و تربیت بمعنای آنست که هر کسی باید در کار تصمیم‌گیری و سیاست‌گذاری‌های آموزشی هرگاه که لازم آید دخالت داشته باشد. اثر پاره‌ای از این تصمیمات بر سرنوشت بخش بزرگی از مردم قابل ملاحظه هست. (برای مثال، امتحانات ورودی دانشگاهها)؛ تأثیرات پاره‌ای فردی است (اراده علی مطالعه دوباره کتاب ریاضی نمی‌باشد) تمهیدات زیادی هست تا معلمان و مربیان باید اتخاذ کنند، تا اینکه به دانش‌آموزان، والدین و نیز عموم مردم یاری رسانند تا آنها بتوانند خود تصمیم گیرنده باشند. مثال‌های زیر نمونه‌هایی از این نوع تصمیمات را نشان می‌دهد:

الف - آیا رضا بایست در یک گروه پیشرفته درسی قرائت قرار گیرد.

ب - آیا مدرسه بایست کتاب انگلیسی اول را تدریس کند و یا کتاب قبلی را و یا اصلاً کتاب جدیدی برگزیند؟

س - آیا در سال اول دوره زبان انگلیسی درس گرامر بقیمت فدا شدن درس تلفظ هنوز مورد توجه است؟

د - آیا من آنطور که باید در کلاس انگلیسی خوب عمل میکنم؟

ی - آیا بایست حتماً به کالج بروم؟

این تنها معدودی از سئوالات بی‌شمار دیگری است که پیش‌روی مربیان، والدین و دانش‌آموزان قرار دارد. زمانی که تصمیمی گرفته می‌شود، سوای اینکه این تصمیم بزرگ یا کوچک باشد، بایست تا حد ممکن براساس اطلاعات کافی و دقیق استوار باشد. هر چقدر این اطلاعات دقیق‌تر باشد تصمیماتی که اتخاذ میشود مطمئناً بهتر خواهد بود. در حقیقت بسیاری از استادانی که امر تصمیم‌گیری را مورد بررسی و مطالعه قرار می‌دهند، معتقدند که یک تصمیم خوب همانی است که بر تمامی اطلاعات مربوط متکی است.

۳-۱- تست، سنجش و ارزیابی

اصطلاحاتی نظیر تست، سنجش و ارزیابی گاهی بجای هم مورد استفاده قرار میگیرند، اما پاره‌ای مربیان بین آنها تفاوت قائل‌اند. واژه «تست» در میان واژگان سه گانه معمولاً دارای بیشترین محدودیت است. «تست» اغلب بر ارائه یک رشته سئوالاتی که می‌باید پاسخ داده شود دلالت دارد. در پی پاسخ‌های داده شده به سئوالات ما معیاری (یعنی معیار نمره دادن) را برای تعیین وضعیت درسی فرد برمی‌گزینیم. معیار «سنجش» اغلب مفهوم گسترده‌تری دارد: ما می‌توانیم وضعیت‌های درسی را با تمهیداتی غیر از امتحانات «تستی» اندازه بگیریم. با استفاده از مشاهده، نمره دادن یا وسایل دیگر که موجب میشود تا ما به شکلی کمی به اطلاعات لازم دست یابیم، سنجش خوانده میشود.

تعریف «ارزیابی» به شیوه‌های مختلف ممکن است طبق نظر استافن بین، (۱۹۷۱) ارزیابی «فرایند طرح و اجرا برای کسب و انتخاب اطلاعات لازم برای مرحله تصمیم‌گیری است. تلقی دوم از مفهوم ارزیابی تعمیق این نکته است که چه نسبیتی بین آنچه که عملاً تدریس شده و اهداف وجود دارد. تعاریف دیگری هم هستند که به سادگی ارزیابی را نوعی فرایند بمنظور داوری حرفه‌ای می‌شناسد که اجازه میدهد تا فرد درباره ارزش و مقبولیت یک معیار به قضاوت بنشیند.

پس ارزیابی نیازمند آنست تا انسان قصد و هدفی در مدنظر داشته باشد. در زمینه آموزش و پرورش ما گاهی به جمع‌آوری اطلاعاتی مبادرت می‌ورزیم که معیاری برای دستیابی به اهداف خاص آموزشی نیستند. اما بدین خاطر لازم می‌آیند که بما کمک می‌کنند تا درباره اهدافی که باید مطرح شوند یا چه شیوه‌های آموزشی باید برگزید تا به اهداف رسید، تصمیم بگیریم. مثالی بزنیم، دو دانش‌آموز ممکن است یک نمره یکسان (نمره امتحانی) دریافت کنند، در حالیکه ما باید تلقی متفاوتی از این دو نمره داشته باشیم. فرض کنید در پایان کلاس سوم مرحله اول دبیرستان (راهنمایی) دو دانش‌آموز داریم که هر دو در پایان دوره در زبان انگلیسی رتبه سوم را کسب کرده است. و حال آنکه یکی از آنها در آغاز دوره رتبه اول و دیگری رتبه دوم را کسب نموده بودند. ارزیابی ما از این دو مورد نبایستی یکی باشد، بدیهی است که یکی از دانش‌آموزان بالاتر از میانگین و دیگری پائین‌تر از میانگین قرار دارد.

ضمناً تذکر این نکته مهم است که بدانیم که خود افراد را سنجش و ارزیابی نمی‌کنیم، بلکه دانسته‌ها و معلومات آنها را در بوته نقد و بررسی می‌گذاریم به میزان تحصیلات دانشگاهی، معلومات انگلیسی، توانمندی در تدریس و از این قبیل. این امر نباید با ارزیابی شخصیت یک فرد اشتباه شود، توجه معلمان، والدین و دانش‌آموزان به این نکته جلب میشود تا بروشنی این وجه تفاوت را بخاطر بسپارند.

۴-۱- آزمون زبان

امتحان گرفتن بخش مهمی از هر برنامه زبان آموزی و یادگیری است. سئوالات خوب و سنجیده، تهیه شده توسط معلمان یا گروهی از زبده‌گان امتحان بروش تست دستکم بدو طریق به دانش‌آموزان کمک میکند.

نخست امتحان مشوق و محرک دانش‌آموزان در یادگیری مطالب درسی می‌باشد. معلمان باید تمامی سعی و تلاش خود را بکار برند تا تجربیات مثبت و پرارزش کلاس از طریق امتحانات تستی بخوبی در اختیار دانش‌آموزان قرار گیرد. پاره‌ای از معلمان گمان می‌کنند کار و وظیفه آنها تنها در یک تدریس بغایت مطلوب خلاصه می‌شود. اما واقعیت آنست که تدریس خوب اگر با سنجش و ارزیابی متناسب همراه نباشد چندان مفید واقع نمی‌شود. یک ارزیابی سنجیده از هر جهت موجبات حس رضامندی در دانش‌آموزان می‌گردد و در خیلی از موارد نارضایتی، شکوه و اعتراض دانش‌آموزان را از برنامه‌های آموزشی تخفیف می‌دهد.

دوم اینکه امتحان به دانش‌آموزان امکان آماده شدن و بدین ترتیب آموختن مطالب درسی را می‌دهد. تکرار برنامه امتحانی موجب خواهد شد تا دانش‌آموزان در زبان مهارت کسب کنند. آنها از نتایج امتحان و بحث پیرامون آن منتفع خواهند شد. در ضمن امتحانات تستی و کوتاه دانش‌آموزان را به اهداف هر رشته درسی و اهمیت آنها واقف خواهد ساخت. تحلیل و بررسی نتایج امتحان موارد مشکلات دانش‌آموزان را روشن و بدین ترتیب فرصت لازم برای جبران ضعف‌ها در اختیار وی خواهد نهاد. بطور کلی این باور وجود دارد که آگاهی بهتری از اهداف کورس و الزامات زبان‌آموزی نزد دانش‌آموزان به ایشان کمک می‌کند تا فعالیت و نیرویشان را در راه رسیدن به اهداف صرف کنند. (مدلن ۱۹۸۳)

۵-۱ آزمون برای چیست؟

معلمان انگلیسی بعنوان زبان دوم / زبان غیرمادری بایستی قادر به توضیح و توجیه فعالیت‌های خود در کلاس باشند. این امر ممکن نیست مگر اینکه بتوانند نتایج آزمون را مورد بررسی قرار دهند. یک امتحان تستی درخور و متناسب باید متضمن پاسخ‌هائی برای سئوالات زیر باشد.

الف) آیا تدریس امری موفق بوده است؟

ب) آیا سطح مطالب درسی متناسب بوده است؟

س) آیا به تمام مهارت‌های زبانی بطور یکسان اهمیت داده شده است؟

د) چه نکاتی لازم به مرور و مطالعه دوباره است؟

ی) آیا همان مطالب در سال بعد می‌باید تدریس شود یا نیازمند تغییر و اصلاحند؟

علاوه بر این، یک بررسی دقیق از نتایج امتحانی ما را قادر می‌سازد تا فرایند ارزشیابی را اصلاح نمائیم. یک چنین بررسی‌ای برای سئوالات زیر پاسخ‌هائی خواهد داشت.

الف) آیا مبحثی که از آن تست بعمل آمده روشن بوده؟

ب) آیا زمان اختصاص یافته کافی بوده؟

س) احساس دانش‌آموزان بهنگام پاسخ‌گوئی به سئوالات چگونه بوده است؟

د) آیا نتایج آزمون بازتاب چگونگی درس خواندن دانش‌آموزان طی تمامی کورس بوده است؟

کاملاً بدیهی است که آزمون‌ها ما را در بهبود وضعیت کیفی تدریس در هر سیستم آموزشی یاری می‌رسانند. هم امتحان شونده و هم امتحان‌کننده از نتایج آزمون منتفع می‌گردند. هیچکسی نمیتواند از اهمیت یک آزمون خوب غافل بماند. یک آزمون خوب حتماً ضرورت ندارد که توسط یک مؤسسه حرفه‌ای تهیه و تنظیم گردد. آزمونی اعم از معلم ساخته یا آماده میتواند در خدمت اهداف مثبت و سازنده باشد. برای روشن شدن موضوع به تفاوت بین آزمون معلم ساخته و آماده در زیر اشاره میکنیم.

۶-۱ آزمون‌های معلم ساخته در مقایسه با آزمون‌های استاندارد^۱

آزمون‌های معلمان غالباً برای ارزیابی پیشرفت درسی دانش‌آموزان در مدارس بعمل می‌آید. در هر سیستم آموزشی، دانش‌آموزان در بوته انواع آزمون‌های تهیه شده توسط معلمان آزموده خواهند شد. ارزش و اعتبار نتایج این آزمون‌ها توسط تمامی دست‌اندرکاران امر آموزش مسئله‌ای پذیرفته شده است. معلمان متعهدند تا به دانش‌آموزانشان بهترین تدریس را ارائه دهند. ضمن انجام آزمون کلاسی آنها میتوانند از طریق تأثیرات آموخته‌هایشان بر دانش‌آموزان کارآئی تدریس خود را هم بسنجند. شوارتز و تیدمان^۲ (۱۹۵۷ صفحه ۱۱۰) معتقدند که آزمون‌های معلم ساخته بارز رشد بدین لحاظ که:

۱- دانش‌آموزان را براساس فعالیت‌های کلاسی می‌سنجند.

۲- به دانش‌آموزان انگیزه میدهند.

۳- فرصتی در اختیار معلمان قرار میدهد تا ضعف‌های دانش‌آموزان را تشخیص دهند.

۴- به معلم کمک میکند تا در صورت لزوم تدریس جبرانی داشته باشد.

علیرغم مفید بودن آزمون‌های معلمان این آزمون‌ها همواره با اعتراضات دانش‌آموزی مواجه میشوند. این شکایات ناشی از ابهامات در محتوای آزمون‌ها و گاهی عدم ارتباط بین آزمون‌ها و مواد آموزشی تدریس شده می‌باشد. ایراد دانش‌آموزان از این قبیل است. «این یک تمرین بی‌فایده بود» «تمیدانم معلم دنبال چی بود» «من تمام بخش‌های عمده کورس را مطالعه کردم اما تنها نکات و توضیحات بی‌اهمیت در امتحان آمده بود» که چندان ایرادات غیرمنتظره‌ای نیست. از این‌رو به هر آزمونی می‌باید بر محتوای پیش‌بینی شده‌ای متکی باشد که معلومات دانش‌آموزان را در زمانی معین بسنجد. چنین آزمون‌هایی می‌تواند توسط معلمان یا گروهی از دست‌اندرکاران حرفه‌ای تهیه و تنظیم گردد.

آزمون‌های معلم ساخته و آزمون‌های آماده در پاره‌ای جهات متفاوتند. به منظور راحتی درک تست‌های استاندارد (همگون شده)، کاملاً عاقلانه است تا اینگونه تست‌ها را مورد بحث و بررسی قرار دهیم. این تست‌ها با هدف تجاری توسط کارشناسان خبره تست ساز و سنجش تهیه و تنظیم میشوند.

۱- استاندارد، یعنی با معیار یا ضابطه‌ای معین سنجیدن، میزان‌دار کردن

آنها با اتکاء به شیوه‌های نمونه‌گیری رفتاری با تمهیدی یکسان به تهیه آزمون مبادرت می‌ورزند. منظور از «تمهیدی یکسان» این است؛ یک رشته سئوالات معین، با هدفی مشخص، طی زمانی معلوم برای نمره دادن، نمره‌گذاری براساس یک روش عینی صورت می‌گیرد، گرچه پاره‌ای از آزمون‌های آماده و استاندارد ممکن است مشتمل بر سئوالات تشریحی باشند. در قیاس با آزمون‌های معلم ساخته، تست‌های استاندارد معمولاً حوزه فراگیرتری دارند (بدین معنی که دروس بیشتری را شامل می‌شوند). این نوع تست‌ها شامل آموزش‌های یکسال تحصیلی یا چیزی بیشتر از این می‌باشند. از طرف دیگر، تست‌های معلم ساخته معمولاً شامل یکدرس یا دروس یک ترم می‌باشد. جدول زیر حاکی از پاره‌ای تفاوت‌های عمده بین آزمون‌های معلمی و تست‌های آماده می‌باشد. (جدول ۱-۱)

جدول ۱-۱: مقایسه آزمون فعلی معلم ساخته و استاندارد (آماده)

ویژگی	آزمون معلم ساخته	آزمون استاندارد
دستورالعمل جهت اجرا و نمره دادن	معمولاً دستورالعمل ثابتی تعیین نمی‌شود	دستورالعمل‌های معین شیوه‌های استاندارد نمره دادن و اجرا
نمونه‌گیری از محتوا	محتوا و نمونه‌گیری هر دو توسط معلم کلاس تعیین می‌گردد	محتوا توسط کارشناسان برنامه تحصیلی و مواد مورد تدریس تعیین می‌شود؛ که شامل تحقیقات گسترده در زمینه برنامه‌های درسی، موضوعات و برنامه‌ها، نمونه‌گیری از محتوا که بطور سیستماتیک انجام گرفته است.
ساختار	ممکن است اتفاقی و شتابزده باشد، اغلب بردن طرح‌های کلی آزمون، بخش گزینش‌ها بخش تحلیل یا تجدیدنظر؛ کیفیت تست ممکن است کاملاً ضعیف باشد.	بکارگیری روش‌های ساخت بسیار دقیق که شامل تدوین اهداف و طرح‌های کلی آزمون، بکارگیری بخش گزینش‌های نهایی، بخش تحلیل سوال و اصلاحات سوال می‌شود
معیارها	فقط معیارهای مربوط به کلاس‌های محدود (محلی) در دسترس است.	علاوه بر معیارهای محدود، آزمون‌های استاندارد بطور نمونه حوزه مدارس ملی و معیارهای ساخته شده توسط مدارس را در دسترس قرار می‌دهد

اهداف و بکارگیری مناسبترین شیوه برای مناسبترین شیوه برای سنجش اهداف
 سنجش مجموعه اهداف برنامه‌ریزی درسی در سطح گسترده و برای
 معین توسط معلم و رقابت‌های بین کلاسی، مدرسه‌ای و کشوری
 رقابت‌های درون کلاسی

۷-۱ آموزش و آزمون زبان

آموزش و آزمودن آنچنان باهم درآمیخته‌اند که کار کردن در یکی از زمینه‌ها بدون دیگری عملاً غیرممکن است. آزمودن از منظر استراتژی آموزش عملی و سازنده به نوآموز فرصت‌های مفیدی برای بحث و گفتگو پیرامون گزینه‌های زبان می‌دهد. همانطور که مدسن^۱ (۱۹۸۳) اظهار می‌کند: آزمون زبان امروزه نشانه نوعی علاقمندی شدید به آموزش و یادگیری جهت تبادل افکار و اطلاعات درست می‌باشد، اما پیش از همه انعکاسی از تمایل به تهیه آزمون‌های معتبر علمی است.»
 در این بخش ما روش‌های آزمون زبان و چگونه تأثیرپذیری آنرا از شیوه‌های عمده آموزش آن در مقاطع مختلف این قرن مورد بحث و بررسی قرار می‌دهیم. ارائه این بخش از آن، بسختی می‌تواند نشان‌دهنده تحولات تاریخی آن باشد.

۷-۱-۱ روش آزمون‌های قدیمی

این آزمون‌ها ارتباط خیلی نزدیکی با شیوه گرامری - ترجمه زبان آموزی دارند. بگونه‌ای ویژه‌تر بگوئیم که این رابطه در مراحل اولیه تحصیلی بسیار قوی‌تر است. این مرحله از آزمون زبان مرحله درک مستقیم است. بسیاری از معلمان با آنکه هیچگاه آموزشی در امر تدریس و آزمون زبان نداشته‌اند، قویاً بر یاد گرفتن زبان و گرامر آن تأکید دارند. در نتیجه دانش‌آموزان ناچارند تا قواعد گرامری بسیاری را بیاموزند و لغت‌های زیادی را حفظ کنند. آنچه که در زیر می‌آید بخشی از همین گونه آیتم‌هاست:

مثال ۱

عبارت زیر را به گذشته برگردانید:

“He drives his car fast.”

مثال ۲

مشتقات اصلی این افعال را بنویسید:

Go, buy, ring, lay

مثال ۳

با استفاده از این کلمات جمله بسازید

Intelligent, bashful, diligent

^۱ - Madsen

مثال ۴ بقارسی برگردانید

A few sentences or a short passage in English.

تست‌های سنتی علاوه بر این شامل مقداری مطلب نوشتنی و درک مطلب بود، این نوع مطالب به آنچه که در آزمون زبان اول مورد استفاده است خیلی شباهت دارد. به نمونه‌های زیر توجه کنید:

مثال ۵

انشاءای بنویسید در حدود ۱۵۰ تا ۲۰۰ کلمه مبنی بر «اهمیت تعلیم و تربیت در کشورهای جهان

سوم»

مثال ۶

متنی برای دیکته

مثال ۷

یک متن خواندنی درک مطلب. ما می‌توانیم انواع مختلفی از سؤالات را پیرامون یک متن خواندنی مطرح سازیم: سؤالاتی مبنی بر تعریف مطلبی پیرامون اطلاعاتی یا درک و استنباط چیزی.

بعد از این مرحله آزمون، (آزمون سنتی)، آزمون زبان به یک مرحله علمی‌تر فرامیرود، ما بسادگی میتوانیم اثرات تعیین‌کننده زبان‌شناسی ساختاری و روانشناسی رفتاری را بر آموزش و هم آزمون زبان در این دوران مشاهده کنیم. یافته‌های این دو مباحث علمی نشان میدهد که در امر مسلط شدن بر زبان بطور علمی و جزء به جزء میتواند مورد سنجش و ارزیابی قرار گیرد. (مدسن ۱۹۸۳)

بعبارت دیگر روانشناسان علم رفتار زبان را مجموعه‌ای از عادات تلقی می‌کنند و در همان حال زبان‌شناسان ساختاری به تحلیل و بررسی اجزاء زبان همچون (اصوات، حروف، واژگان، و ترکیب) میپردازند. یکی دیگر از دلایل توسعه (فراگیر شدن) آزمون‌های تستی ضریب بالای عدم دقت در آزمون‌های نظری بوده است. این درست است که همه ما ارزش متن نویسی در تمرین انگلیسی را درک میکنیم، اما خیلی از معلمان نسبت به مدل ارزیابی نظری، نظر خوشی نداشتند. در این جا اشاره کنیم که بهر حال بسیاری از آزمون‌های گرامری سنتی نیز تا آنجا که به نمره دادن مربوط میشود عینی به حساب می‌آیند. بگذارید به نمونه‌های آزمونی زیر نگاهی بیافکنیم و آنگاه سؤالات در رابطه با آنها را مورد بحث و بررسی قرار دهیم.

مثال ۸

جملات زیر را کامل کنید، بیشتر از پنج کلمه استفاده نکنید:

"She didn't go to school because"

مثال ۹

جمله زیر را سوالی کنید.

"She Works in a library."

مثال ۱۰

شکل درست فعل را در پرانتز بنویسید.

He (go) to school before you came.

مثال ۱۱

شکل درست فعل را برگزینید.

If you need it , he it to you.

a) Lends

c) will lend

b) to lend

b) has lent

مثال ۱۲

جمله درست را برگزینید

a) He at school studies English.

b) English at school he studies.

c) He studies at school English.

d) He studies English at school.

مثال‌های فوق دو گونه آزمون عینی (Objective) را نشان می‌دهند. واضح است که یکی از آن میان از دیگری عینی‌تر است. در مثال ۸ دانش‌آموزان با جمله‌ای که پایان آن باز و کامل نشده است مواجه‌اند. بهمین خاطر ممتحن به مشکل نمره دادن و سنجش برمی‌خورد. در مثال ۱۲ دانش‌آموزان سه گزینه شکل نادرست در اختیار دارند. ارائه گزینه‌های غلط به دانش‌آموزان یک مسئله مورد بحث و جدل است. چنین گمان می‌رود که مواجه شدن مرتب دانش‌آموزان با اشکال غلط زبانی منجر به تثبیت این اشکال غلط در ذهن دانش‌آموزان خواهد شد. بعدها چگونه می‌شود این‌ها را تصحیح کرد؟ شکل ظاهری پاره‌ای از آزمون‌های عینی ناتمام و پاره‌ای چند گزینه‌ای هستند. در این بخش ما تنها به نوع دوم خواهیم پرداخت.

۷-۱-۲ تست‌های چند گزینه‌ای

تست‌های چند گزینه‌ای از متداول‌ترین گونه آزمون‌های عینی هستند. دانش‌آموزان معمولاً با سه، چهار یا گاهی پنج راه یا گزینه مواجه هستند. (یک مورد درست و بقیه انحرافی). از دانش‌آموز انتظار می‌رود تا جواب درست را از میان چند پاسخ غلط بیابد. کار دانش‌آموزان کاملاً ساده و نمره دادن بمراتب ساده‌تر است. به نمونه زیر توجه کنید:

مثال ۱۳

He left home Ten o'clock

a) in

b) at

c) by

d) until

این شکل نمادین یک تست گرامری است. گفته می‌شود که این چنین مواردی تنها نکته‌ای از نکات مختلف گرامری را مورد توجه قرار می‌دهد.

ما تصور می‌کنیم که آزمون‌های چند گزینه‌ای دانش‌آموزان را با متنی محدود مواجه می‌سازد که همان صورت سؤال است و از آن فراتر نمی‌رود. امتحان‌دهنده (آزمون شونده) اطلاعاتی اضافی درباره

گوینده یا وضعیت دریافت نمیکند و در نتیجه انتخاب گزینه درست مشکل است. در واقع بسیاری از قواعد گرامری در شرایط معمولی رعایت نمیشوند. بدین ترتیب گمان میرود که بشود برای خیلی از این گزینه‌های انحرافی کاربرد عینی پیدا کرد.

خلاصه اینکه گفته میشود که سؤالات خوب یک آزمون سازنده همراه با گزینه‌های انحرافی منطقی کاری است بس مشکل، و متأسفانه خیلی از آزمون‌های معلمان آزمون‌های بد و نامناسبی هستند بخشی از توصیه الیزابت اینگرام^۱ را در ذیل میاوریم.

«آزمون سازان تازه کار نخست باید از گزینه جملات ناتمام شروع کنند. این گزینه‌ها بایست در اختیار دانش‌آموزانی از آن نوع قرار گیرد که معلم بالاخره میخواهد از ایشان امتحان بگیرد. پاسخ‌های غلط میتواند بعنوان گزینه‌های انحرافی بعدها در آزمون بکار آید.»

ضمناً باید اشاره کرد که استفاده از گزینه‌های ناموجه و غیرمنتظره در آزمون زبان، شیوه‌ای اشتباه است. غالباً بسیاری از این آزمون‌ها گزینه‌هایی را پیش چشم دانش‌آموزان میگذارند که انگار هم‌اش اشتباه است در حالیکه در واقع اشتباه نیست، زبانی که هم‌اش غلط است جز یکی. به نمونه‌های زیر توجه کنید.

مثال ۱۴

The teacher me what to do

- a) told b) suggested c) explained to d) said

مثال ۱۵

How is everything?

- a) Not so good b) all right
c) thank you so much d) is all right

مثال‌های ارائه شده به تمامی جملات بی‌ربطند. گذشته از این میشود متن‌های بلندتری تنظیم کرد که حاوی سؤالات و گزینه‌های بیشتری باشند. همانطور که قبلاً اشاره شد، رفتارگرایان به فرایند زبان‌آموزی بعنوان یادگیری یک سلسله عادات مینگرند. این شبکه پیچیده عادات گام به گام شکل میگیرند. بعلاوه بسیاری از آزمون‌های تهیه شده توسط معلمان برآند تا عناصر زبانشناسی را جدا جدا بازمایند. بدین گمان که همین‌ها مهارت‌های زبانی را شکل میدهند. پس بدین ترتیب می‌شود به تنظیم و تهیه آزمون‌هایی همت گذاشت که چهار مهارت سنتی زبان را به‌همراه عناصر مرتبط با آن بسنجد. بسیاری از این آزمون‌های مربوط به عناصر مجرد زبان همانطور که گفته شد از نوع چند گزینه‌ای هستند. بحث مفصل‌تری پیرامون تهیه این‌گونه آزمون‌ها و سنجش مهارت‌های مختلف در فصول بعد ارائه خواهدشد.

^۱ - Elizabeth Ingram

در مراحل بعدی آزمون زبان پیشرفته، تعدادی از آزمون‌ها بسمت آزمون‌های متداول جهانی پیش می‌روند. بسخن دیگر، این‌گونه آزمون‌ها از زبان آموزان درخواست‌های جامع‌تری دارند. دو نوع مهم از این آزمون‌های جهانی شده تست دیکته و کلوز می‌باشد. در همینجا باید بگوئیم استفاده از دیکته همیشه محل بحث و جدل بوده است. از طرف دیگر تست کلوز^۱ (برگرفته از روانشناسی گشتالت) مبتنی است بر متنی که کلماتی از آن حذف گردیده‌اند. یافتن این کلمات نیازمند مهارت‌های بیانی و اطلاعات کافی از واژگان زبان و قواعد گرامری است. علاوه بر سود بردن از علائم زبان شناسی به دانش آموز برای پاسخ دادن به چنین تستی باید بر پاره‌ای نشانی‌های دیگر متن متکی باشد.

۳-۷-۱ آزمون ارتباطات (زبان مکالمه)

زبان آموزی بطور روزافزونی با امر ارتباطات پیوند یافته و اهداف آن مورد ارزیابی مجدد قرار گرفته است. آزمون زبان در هر حال تا حدودی نتوانست تکنیک‌های مؤثر ارزیابی و سنجش استفاده از زبان در شرایط مکالمه را تحول بخشد. در بسیاری از روش‌های مورد استفاده ترجیح داده می‌شود تا توانمندی دانش‌آموزان برای کارآمدی در سیستم گرامری و فونولوژی زبان سنجیده شود. با این همه میتوان آزمون‌هایی تهیه کرد که دانش‌آموزان را قادر سازد تا زبان کاربردی را بیاموزند، و اظهاراتی را که بلحاظ اختصاصات خود به کاربرد معینی از زبان مربوط میشوند تشخیص دهند. این موج که به «گرایش کاربردی زبان» معروف است بعداً مورد بحث و بررسی قرار خواهد گرفت.

زبان در واقع چیزی بیشتر از ردیف کردن واژگان است. زبان درست و واقعی چیزی است فراتر از کلمات، قواعد گرامری یا تنظیم بصورت جمع. سوء تفاهمی عمومیت یافته درباره زبان اینست که کاربرد موفقیت آمیز زبان به برخورداری موفقیت آمیز از امکانات آن میانجامد. در حالیکه اصلاً چنین نیست. زیرا که جنبه‌های زبانشناسی زبان تنها بخشی از فرایند ارتباطی است. برای مثال عبارات زبان در قالب یک متن اجتماعی معنی‌ای دارد که با آنچه بطور مجرد در جمله بکار میرود متفاوت است. مثال زیر این نکته را روشن خواهد ساخت.

صحنه‌ای را تصور کنید که شما با یک بغل کتاب، قصد خروج از اطاق را دارید. درب کلاس بسته است و شما نیاز به کمک دارید. بنابراین خطاب به یکی از دانش‌آموزان می‌گوئید:

Do you think you could open that door?

دانش‌آموز (بدون آنکه برای باز کردن درب حرکتی کند) می‌گوید:

"Yes I do."

این دانش‌آموز بدانگونه که شما انتظار داشته‌اید پاسخ نگفته است. از این مثال چنین برمیآید که شاید معنی را خوب نرسانده باشید اما مسئله نه استفاده نادرست از کلمه و نه اشکال گرامری است. ما همه میدانیم که منظور از جملات سئوالی مجرد بلحاظ گرامری کسب اطلاعات و جواب Yes یا No

^۱ . close

است. همان جوابی که دانش‌آموز بدرستی در پاسخ شما داده است. با این وجود همه میدانیم که جملات سوالی در واقع یکی از رایج‌ترین روش‌های درخواست محترمانه برای انجام کاری است. در خاتمه، در حال حاضر ادبیات TEFi حاکی از آن است که آموزش زبان انگلیسی بمنظور مکالمه و اغلب مکالمه تخصصی متفاوت، از گرایشات سنتی‌تر است، اما در ارزیابی توانمندی مکالمه دانش‌آموزانمان این تست چقدر درست عمل می‌کند؟ پاره‌ای از استادان در این زمینه فکر میکنند که این تست از قافله عقب مانده است.

فعالیت‌ها

- ۱- چه پیشنهادی را در آزمون توانمندی زبان دانش‌آموزان بیشتر از همه مفید میدانید؟
- ۲- به چه میزانی معلمان شما بهنگام آزمون زبان از شیوه رفتاری استفاده میکنند؟
- ۳- خوبی و بدیهای انشاء نوشتن بمنزله یک امتحان چیست؟
- ۴- اشاره به این نکته که اکثر معلمان از دیکته نوشتن بعنوان راهی برای آزمون درستی دیکته کلمات استفاده میکنند چقدر است؟

۵- متن زیر و مکالمه بین آقای احمدی و آقای هاچ درصدد شغل معلمی است بخوانید. سپس به سئوالات پاسخ دهید. نظر به اینکه هر پاسخی ممکن است به آنها از ۱ تا ۳ نمره بدهید. به گزینه نخستین نمره ۳ بدهید و الی آخر.

آقای احمدی: صبح بخیر آقای هاچ بفرمائید بنشینید.

آقای هاچ: ۱- صبح بخیر آقای احمدی متشکرم

۲- روز خوبی برایتان باشد آقا

۳- صبح بخیر، خیلی متشکرم

آقای احمدی:

درخواستتان روی میز من است. اما خیلی دوست دارم راجع به تجربیاتتان چند سؤال بپرسم.

آقای هاچ: ۱- چرا که نه؟ بفرمائید.

۲- خواهش میکنم، تردید نکنید. خیلی خوشحال بشوم که جواب بدهم.

۳- (پاسخ نمیدهد با نوعی عصبانیت لبخند میزند.)

آقای احمدی: باخبر هستم که شما چند کلاس درس EFL در نوبت عصر دارید.

آقای هاچ:

۱- درست است. دو دسته کلاس دارم. یکی کلاس کودکان و یکی هم نوجوانان

۲- بله واقعاً همینطور است. کاملاً درست میگوئید. من مسئولیت تدریس به دو گروه در

سطوح مختلف را بعهده دارم.

۳- بله همینطور است. خیلی هم برایم جالب است. دو نوبت عصر در هفته را در مؤسسه تدریس دارم. یک کلاس مقدماتی در پنجشنبه‌ها و یکی کاملاً پیشرفته در روز سه‌شنبه.

آقای احمدی: کدام پایه را ترجیح می‌دهید؟

آقای هاچ: ۱- خوب، فرقی نمیکند. فرق میکند؟

۲- تدریس پیشرفته را قدری مشکلتر و در عین حال تجربه آنرا مفیدتر میدانم.

۳- قلباً باور دارم که به تدریس من در مرحله مقدماتی بیشتر نیاز هست.

الف) این گزینه‌های سه‌گانه را چگونه نمره‌گذاری میکنید؟ مثلاً اگر یکی درست یا کاملاً مناسب باشد آنهای دیگر چگونه‌اند؟

ب) بنظر شما اگر آقای هاچ مرتب تکرار میکرد. الف) گزینه اول شما

ب) گزینه دوم شما س) گزینه سوم شما چه تأثیری بر مخاطب می‌گذاشت؟

س) بعضی از مشخصه‌های عادی گزینه‌های سوم شما چیست؟

ط) آیا شرایطی را که منجر به گزینه‌های مناسب دو و سه شده را میتوانید تصور کنید؟

ی) یکی از گزینه‌های در گفتگوی آقای هاچ بزبان هیچ نگفتن است. یک پاسخ منفی. یک چنین

پاسخی چگونه توجیهی دارد؟

فصل دوم

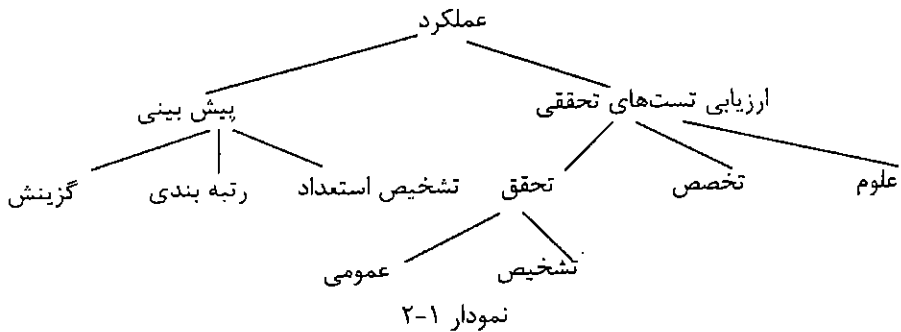
نقش آزمون زبان

۲-۱ مقدمه

همانطور که قبلاً اشاره شد، آزمون تستی روشی است برای جمع‌آوری اطلاعات عددی بر مبنی برگه امتحانی. هدف از جمع‌آوری اطلاعات عبارت است از اندازه‌گیری موارد درست در یک امتحان. ماهیت امتحانی که از طریق آن اطلاعات لازم با هدف ارزیابی جمع‌آوری میشود و ماهیت درستی که این سنجش بخاطرش انجام میگردد، بایستی بروشنی تصریح گردد. این امر مستلزم آنست که پیوندی تنگاتنگ و در پاره‌ای موارد نکته به نکته بین امتحان و درس موردنظر وجود داشته باشد. برای تامین یک چنین پیوندی، هدف از امتحان و نوع تصمیماتی که بر مبنای نمرات داده شده گرفته می‌شود باید بروشنی معین گردد. این فصل به هدف و کار و وظیفه یک آزمون میپردازد.

وظیفه یک آزمون به هدفی که برای آن طرح و تنظیم میشود مربوط است. ممتحن بایست هدف از آزمون را بروشنی دریافته باشد. والا استفاده از آزمون با هدفی ناشناخته منجر به اتخاذ تصمیماتی غیرموجه و لاجرم نتایج نامطلوب خواهدشد. از آنجائیکه آزمون زبان از اصول و قواعد سنجش در آموزش و پرورش تبعیت میکند، بهمین خاطر هم اصول تئوریک و عملی حاکم بر امر سنجش و ارزیابی آموزشی بایستی در آزمون زبان مدنظر قرار گیرد. مطابق با این اصول، آزمون‌ها دو وظیفه اساسی برعهده دارند؛ برآورد نتایج (پیش‌بینی) و ارزیابی آنها. آزمون‌های تخصصی، که عبارتند از تست‌هایی برای گزینش شدن، انتخاب جا و تشخیص استعداد مستقیماً ربطی به کورس خاص درسی ندارند. سنجش تست‌های attainment از طرف دیگر بر کورس‌های کاملاً مشخص درسی مبتنی هستند و متضمن آزمون‌های کار و مهارت و معلومات می‌باشند.

این طبقه‌بندی در نمودار شکل ۲-۱ به تفصیل مورد بحث قرار خواهد گرفت.



۲-۲ تست‌های پیش‌بین

واژه «Prognostic» بمعنی پیش‌بین است. آزمون‌های پیشرفته بمنظور پیش‌بینی کورس‌های درسی آینده دانش‌آموز مصرف دارند. نمرات تست‌های پیش‌بین در راستای تصمیم‌گیری درباره فرصت‌های مناسب تحصیلی و شغلی آزمون‌شونده مورد استفاده قرار می‌گیرد. بویژه اینکه آزمون‌های پیش‌بین مسئولان را قادر می‌سازد تا راه حل مشکلات مربوط به مقاطع اساسی و مناسب درسی یادگیری درست زبان‌های خارجی، و اشتغال آینده دانش‌آموزان را بیابند. سئوالات مربوط به اقدامات آینده در خصوص دانش‌آموزان نشان‌دهنده آنست که آزمون‌های پیش‌بین ارتباطی به زمینه‌های تحصیلی آنان هرگز ندارند. محور اصلی این‌گونه آزمون برای تصمیم‌گیری درست درباره موفقیت آینده آزمون‌شونده بر مبنای شایستگی‌های امروزی اوست. طبقه‌بندی عمده تست‌های پیش‌بین یعنی تست گزینه‌ها، انتخاب جا و تست استعداد در مباحث زیر مورد بررسی قرار گرفته است.

۲-۲-۱ آزمون‌هایی برای گزینه‌ها

منظور از این‌گونه تست‌ها تهیه اطلاعاتی است که براساس آنها بتوان نسبت به پذیرش یا عدم پذیرش آزمون‌شونده در یک برنامه خاص تصمیم گرفت. این برنامه الزاماً به حوزه‌های درسی و آموزشی ربط پیدا نمی‌کند. بلکه به تمامی مؤسساتی که بنوعی با گزینه‌ها یا عدم گزینه‌ها گروهی از متقاضیان سروکار دارند مربوط می‌شود. مثلاً، برای دریافت گواهی‌نامه رانندگی از داوطلب یک تست گزینه‌ها گرفته می‌شود مبنی بر اینکه قبول یا احتمالاً رد خواهد شد. معیار برای قبول شدن یا رد شدن در دست مسئولان است. نمونه دیگری از این تست‌های گزینه‌ها مربوط است به گزینه‌ها افراد بخاطر استخدام. از متقاضی معمولاً یک تست برای گزینه‌ها گرفته می‌شود تا معلوم شود میتواند استخدام گردد یا نه. مثلاً برای اینکه معلوم شود یک ماشین‌نویس حائز شرایط گرفتن پروانه هست یا نه از او میخواهند

نمونه‌ای را تایپ کند. اگر طبق معیار ۲۰۰ کلمه در دقیقه را با حداکثر سه اشتباه بتواند تایپ کند آنگاه به دریافت گواهی‌نامه نائل می‌آید. در محیط‌های آموزشی نیز تست‌های گزینشی در شرایط متفاوتی مورد استفاده قرار می‌گیرد. برای نمونه کورس‌های Non-credit براساس قبول یا رد برنامه‌ریزی می‌شوند. آنهایی که نمره از قبل معیار قرار داده شده را اکتساب می‌کنند قبول و بقیه مردود شناخته می‌شوند. در واقع، سیستم آموزشی در ایران براساس روش قبولی - مردودی و معیار ۵۰ درصد استوار است.

یک فاکتور مهم در استفاده از تست‌های گزینشی این است که زمانی که سقف نمره از قبل مشخص شد، دیگر محدودیتی در تعداد قبولی‌های آزمون اعمال نمی‌شود. گرچه به اقتضای شرایط واقعی، و بلحاظ ملاحظه است و محدودیت‌های اداری پذیرش تمامی کسانی که آزمون را با نمره قبولی گذرانده‌اند ممکن نیست. به همین دلیل جدای از بعضی موارد، نظیر اعلام نتایج و صدور گواهی‌نامه در بسیاری از موارد دیگر، تعداد بی‌شمار متقاضیان حتی قبولی‌های با نمره بالا، امکان پذیرفته شدن نمی‌یابند. بدین معنی که تعداد متقاضیانی که از امتحان می‌گذرند از ظرفیت برنامه آموزشی بالاتر است. در چنین صورتی تست گزینش، یک آزمون رقابتی است.

در چهارچوب آزمون رقابتی نمرات متقاضیان از بالاترین تا پایین‌ترین، رتبه‌بندی می‌شوند. سپس گزینش به همین منوال از بالا به پایین به آخرین نفری که پذیرفته می‌شود سیر می‌کند. داوطلبانی هستند که بعد از آخرین نفر پذیرفته شده قرار دارند و ممکن است حائز شرایط و حتی بالاتر از معیار تعیین شده قرار داشته باشند. اما بلحاظ عدم امکانات مؤسسه آموزشی در هر صورت نتوانند پذیرش بگیرند. در چنین وضعیتی این کار برای متقاضیان بیشتر نوعی مسابقه و رقابت است تا قبول شدن.

برای اجتناب از چنین محدودیت‌هایی، دو گزینه کلاً ممکن است. گزینه اول توسعه امکانات بمنظور جذب متقاضیان بیشتر است. گزینه دوم تعدیل معیار موفقیت در امتحان می‌باشد. اگر متقاضیان در زمینه معینی صاحب معلومات بالایی هستند، خوب می‌شود سطح آزمون را بالاتر گرفت. در وضعیت قبلی، امکان پذیرفته شدن تعداد بیشتری از متقاضیان افزایش خواهد یافت. در حالیکه در شرایط دوم، تعداد درخواست‌کنندگان که از خان امتحان می‌گذرند شمار محدودتری خواهد بود که امکانات موجود به ایشان خواهد رسید.

یکی از نمونه‌های خوب رقابت در امتحانات ورودی دانشگاه‌های ایران قابل مشاهده است. اگر دانشگاه‌ها قادر به پذیرش تمامی داوطلبان که از امتحان موفق بیرون می‌آیند باشند در آن صورت امتحانات دیگر رقابتی نیست، گزینشی است. معهذا، بلحاظ شمار فوق‌العاده زیاد داوطلبان روش‌های گزینشی به روش‌های رقابتی بدل گشته است. از طریق آزمون رقابتی، داوطلبان تحصیل از میان بالاترین رتبه، تا آخرین نفری که دانشگاه ظرفیت پذیرش دارد گزینش می‌شوند. نمره آخرین فرد، هر چه که می‌خواهد باشد می‌تواند نمره قبولی به حساب آید. نتیجه زحمت‌افزای روش‌های رقابتی نوسانات معیار قبولی و مردودی است. با توجه به موفقیت سراسری آزمون شوندگان و ظرفیت دانشگاه‌ها فردی با

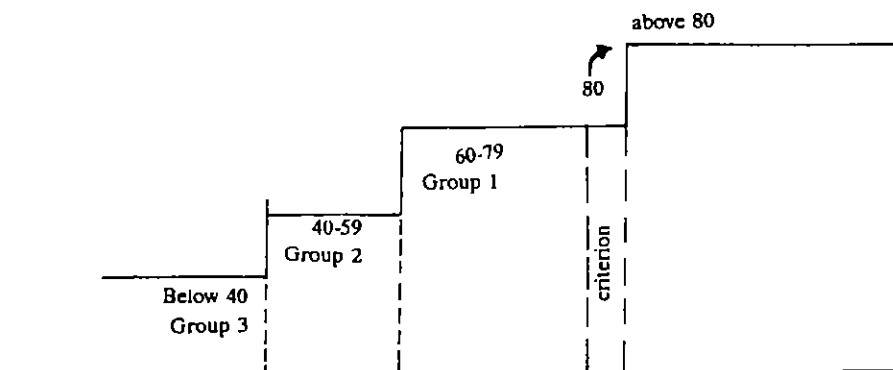
همان درصد انتخاب گزینه‌ها ممکن است یکجا نمره قبولی بگیرد و جای دیگر نگیرد. همانطور که اشاره شد، یکی از راه‌های مرتفع شدن مشکل، تعدیل معیارهاست. یعنی سطح نمرات را بالا یا پائین بیاورند. امکان دیگر افزایش ظرفیت مؤسسات آموزشی برای پذیرش‌های قبول شدگان آزمون سراسری می‌باشد.

۲-۲-۲ آزمون‌های رتبه‌بندی (تعیین سطح)

از این نوع آزمون‌ها بمنظور تعیین مناسبترین مقطع یادگیری برای آزمون شونده استفاده میشود. برعکس آزمون‌های گزینشی، در آزمون‌های رتبه‌بندی قبولی، مردودی وجود ندارد هدف از رتبه‌بندی صرفاً اندازه‌گیری ظرفیت‌های یک متقاضی در ادامه مسیر معین بهنگام زبان‌آموزی است. کمک رتبه‌بندی مناسب و آموزش کافی، به تمامی متقاضیان باید کمک شود تا به سطح معینی از توانایی زبان مورد لزوم برای دستیابی به معیار موردنظر برسند.

آزمون‌های کلاس‌بندی یکی از متداول‌ترین تست‌ها در سطوح مختلف آموزش زبان می‌باشد. بعنوان مثال، تصور کنید دانش‌جویان بخش زبان دانشگاه برای ادامه رشته دانشگاهی آشنائی کافی به زبان انگلیسی الزامی باشد. امکان اینکه هر دانش‌جوی پذیرفته شده مهارت لازم را در زبان نداشته باشد. بنابراین، بخش زبان آزمون رتبه‌بندی را مقرر می‌سازد. بعلاوه تصور کنید که برحسب داده‌های تجربی بدست آمده در بخش زبان، اکتساب نمره ۸۰ در آزمون حاکی از امکان ادامه تحصیل در دانشگاه باشد. به سخن دیگر دانش‌جویی که نمره ۸۰، یا بالاتر را دریافت میکند دیگر لزومی به شرکت در کلاسهای زبان فوق‌العاده ندارد. این بدین معناست که سایرین که نمره اکتسابی‌شان زیر ۸۰ می‌باشد می‌باید حتماً برای ارتقاء مهارت زبانی در کلاس‌های تقویتی شرکت کنند.

چنین دانش‌جویانی به روش‌های مختلف میتوانند کلاس‌بندی شوند. برحسب تصمیمات بخش زبان، تنها در سه گروه طبقه‌بندی میگردند: دریافت‌کنندگان نمرات ۶۰ تا ۷۹؛ ۴۰ تا ۵۹ و نمرات زیر ۴۰ این تقسیم‌بندی در نمودار زیر آمده است.



شکل ۲-۲: طبقه‌بندی دانشجویان

همانطور که شکل ۲-۲ نشان میدهد دانشجویان رتبه ۳ خیلی پائین‌تر از معیار لازم هستند و بنابراین برای رسیدن به سطح مورد لزوم بخش می‌باید تحت آموزش فوق‌العاده قرار بگیرند. در مقایسه، دانشجویان گروه ۲ به نسبت گروه ۳ نیاز به آموزش کمتری دارند. به بیان دیگر، هر چقدر این‌گونه دانشجویان از معیار دورتر باشند بهمان اندازه نیاز بیشتری به آموزش خواهند داشت. بدین‌سان بدیهی است که سیاست غلطی خواهد بود در صورتی که آزمون رتبه‌بندی بگذاریم و دانشجویان را کلاسبندی کنیم. اما مطالب درسی و آموزش یکسانی به همه آنها بدهیم. در حالیکه هدف از روش رتبه‌بندی کمک به آنهایی است که نیازمند آموزش بیشتری هستند. بنابراین، سطوح متفاوت توانمندی آنها نخستین اهمیت را در امر آموزش مطالب دارد. هدف آموزش در این سطوح می‌باید بر اصل رفع تفاوت‌های گروهی و کمک به دانشجویان برای رسیدن به سطح لازم باشد. چنانچه در این زمینه بدون توجه به تفاوت‌ها از آموزش یکسانی برخوردار گردند نه فقط تفاوت‌ها سر جای خود باقی خواهند ماند. بلکه آزمون هم به هدف نخواهد رسید.

بمنظور تحقق این اهداف، دو روش متفاوت و در مقابل همدیگر، الزاماً انحصاری میشود بکاربرد. روش نخست ناظر بر طول مدت آموزش است. آنهایی که در مرحله مقدماتی هستند بایستی به نسبت متوسط‌ها و پیشرفته‌ها زمان بیشتری آموزش ببینند. برای مثال گروه مقدماتی بمدت سه ترم تحت آموزش قرار میگیرند در حالیکه این مدت برای گروه متوسط‌ها دو ترم، و پیشرفته‌ها تنها یک ترم آموزش می‌بینند.

روش دوم به کمیت ساعات آموزش مربوط میشود. بدین‌معنی مقدماتی‌ها به نسبت متوسط‌ها یا پیشرفته‌ها ساعات بیشتری آموزش می‌بینند. برای مثال گروه مقدماتی ۱۰ ساعت در هفته برنامه آموزشی دارند در حالیکه متوسط‌ها ۶ ساعت و پیشرفته‌ها بمدت ۳ ساعت.

متناسب با سیستم آموزشی، یکی از شیوه‌های فوق، یا هر دو آنها در یک‌زمان، باید مورد استفاده قرار گیرد. در هر حال بلحاظ پاره‌ای محدودیت‌های زمانی، آموزش فشرده از آموزش طولانی مناسب‌تر است. بدون توجه به شیوه مورد استفاده، تفاوت بین معلومات دانشجویان در این زمینه می‌باید تا سر حد امکان مرتفع گردد. والا آزمون رتبه‌بندی، که منظور از آن آماده ساختن دانشجویان بخاطر مطالعه بیشتر می‌باشد به هدف نخواهد رسید.

۳-۲-۲ تست‌های تشخیص استعداد

آزمون‌های استعداد بر مبنای پیش‌بینی موفقیت متقاضیان در دستیابی به پاره‌ای اهداف در آینده می‌باشد. برای آنکه آزمون استعداد بعمل آید تست شونده نیازی به داشتن اطلاعات اولیه از مطلبی که قرار است مورد آزمون قرار گیرد ندارد. در آموزش زبان، و از طریق تست استعداد، میشود تشخیص داد که یک نفر کدام زبانی را میتواند ساده‌تر و با موفقیت بیشتر بیاموزد.

تست‌های استعداد میتوانند منجر به اتخاذ تصمیم در مورد وضعیت اشتغال متقاضیان در آینده کارشان نیز گردد. این تست‌ها عملکرد وسیعی دارند، از تشخیص حوزه‌های عمده درسی دانش‌آموز گرفته تا پیش‌بینی اشتغال آینده متقاضیان و پیشنهاد مبنی بر یادگیری یک زبان غیر از زبان مادری از اهم وظایف آنهاست. مثلاً بکمک نمرات تست استعداد پیش‌بینی میزان مسئولیت‌های معینی که فرد در آینده تعهد خواهد کرد. نظیر یک مهندس، پزشک، وکیل یا معلم تا چه حد میتواند خوب عمل کند.

البته، بسط و توسعه آزمون‌های استعداد کاری است ظریف و وقت‌گیر. قبل از هر پیشنهادی بر مبنای نمرات تست، میزان برد تخمینی آنها می‌باید بطور تجربی برآورد گردد. تنها پس از چند بار تمرین و تجربه فرد میتواند براساس تست‌های استعداد اقدام به پیش‌بینی صحیح نماید. والا استفاده از تست‌های متفرقه unestablished منجر به کسب اطلاعات غلط و گاهی گمراه کننده درباره فرد مورد آزمون می‌شود.

۲-۳ نمره‌گذاری تست‌های تحقیقی:

دومین طبقه‌بندی عمده بلحاظ کارکرد آزمون‌ها مشتمل بر تست‌هایی است که هدف از آنها ارزیابی آموخته‌های فرد است. برخلاف تست‌های پیش‌بین، معیار این‌گونه تست‌ها اندازه‌گیری میزان آموخته مطالبی است که به فرد ارائه شده است. این امر بمعنی آن است که این تست‌ها نسبت به تست‌های پیش‌بین بطور مستقیم‌تری با سیستم آموزشی پیوند دارند. انواع این تست‌ها شامل تست آموخته‌ها، تست مهارت و تست معلومات می‌باشند.

۲-۳-۱ تست‌های تحقیقی

منظور از طرح و تنظیم این نوع تست‌ها اندازه‌گیری میزان یادگیری دانش‌آموزان در چهارچوب یک سری مطالب درسی است. این‌گونه تست‌ها نقشی مهم در محیط‌های آموزش بلحاظ آنکه اکثر تست‌های کلاس در همین رده قرار دارند بازی میکنند. امتحانات میان ترم، پایان ترم و کلاس از جمله معدود تست آموخته‌ها به حساب می‌آیند.

بسیاری از تست‌های تحقیقی اغلب توسط معلمان کلاس تنظیم می‌شوند. زیرا که این تست‌ها بایستی بر مطالبی متکی باشد که در کلاس تدریس می‌گردند. البته تست‌های آماده‌ای هم هستند که توسط مؤسسات حرفه‌ای تهیه میشوند. بیشتر این تست‌ها معلومات تست سونده‌ای را که قرار است طی یک دوره یا چند دوره بیاندوزد اندازه می‌گیرد. یک چنین تست‌هایی «تست آموخته‌های عمومی» خوانده میشوند. مثلاً، تست‌های سال اول راهنمایی، تست‌های سه سال و هفت سال دبیرستان جزء تست‌های عمومی هستند.

بهر حال، معلمان و مدیران گاهی تصمیم می‌گیرند تا میزان معلومات دانش‌آموزی را براساس موضوع خاص بسنجند. این تست‌ها، که بطور ویژه و هدفدار جزئیات توضیحی موضوعات آموخته شده را

میسنجند، تست‌های تشخیص خواننده میشوند. تست‌های تشخیصی برای تعیین نقاط ضعف و قوت یک آزمون شونده در یک کورس خاص درسی بدور از هرگونه ابهام عمل میکنند. برای مثال آزمون زبان انگلیسی‌ای که بر مبنای مطالب تدریس شده و بمنظور اندازه‌گیری معلومات کلی در کلاس زبان انگلیسی تنظیم یافته تست عمومی آزمون درس انگلیسی به حساب میاید. یک چنین تست‌هایی مشتمل بر تست‌های فرعی ساختار، واژگان، درک مطلب و دیگر مباحث مربوط به مواد ددرسی ضمن دوره را شامل می‌شود. از سوی دیگر تستی که برای بیان قوت و ضعف دانش‌آموزان در حوزه‌های ویژه مثل ساختار زبان تهیه شده همینطور تست تشخیص تلقی میگردد.

معلمان و مربیان از طریق تست‌های تحقق برنامه خاصه تست‌های تشخیصی آن میزان از کمک‌های خود به دانش‌آموزان را که باعث تحقق اهداف برنامه شده است اندازه میگیرند. با جمع‌آوری اطلاعات مربوط به ضعف و قوت دانش‌آموزان، معلمان روش‌های آموزشی را بگونه‌ای تعدیل یا اصلاح میکنند که با توجه به نقاط ضعف دانش‌آموزان به اهداف برنامه برسند. آنگاه این تست‌ها میتوانند هم برای اهداف آموزشی و درسی و هم سنجش و اندازه‌گیری بکار روند.

یکی از مزیت‌های عمده و عملی تست‌های تحقق برنامه بر سایر تست‌ها مفید بودن آنها قبل از آغاز هر برنامه آموزشی است. از طریق آزمون‌های تشخیصی، و بمنظور برآورد مشکلات مبرم دانش‌آموزان معلمان اقدام به جمع‌آوری اطلاعات مینمایند. آنگاه براساس چنین اطلاعاتی، آنها به تنظیم فعالیت‌های آموزش‌ای مبادرت میورزند که سرانجام به رفع این مشکلات میانجامد، ضمناً معلمان میتوانند برنامه‌های آموزشی را بمنظور پاسخ به نیاز دانش‌آموزان از طریق تشخیص پارامترهایی که می‌باید در کلاس مورد توجه قرار دهند تهیه و تنظیم کنند.

۲-۳-۲ تست‌های تخصصی

برخلاف تست‌های تحقیقی، که بر اندازه‌گیری آنچه که دانش‌جو ضمن مطالب درسی یک کورس دانشگاهی بدست آورده است متمرکز می‌باشد، این گونه تست‌ها با هدف سنجش میزان توانائی کلی دانش‌جویان در زمینه زبان با توجه به موارد زیر تنظیم میگردد: الف) میزان معلوماتی که دانش‌آموز ضمن آموزش زبان اندوخته است. ب) میزان توانمندی وی در اجزاء ساختاری زبان (س) مهارت‌های زبانی وی. در تست‌های سنجش مهارتی راههایی که دانش‌جویان به فراگیری زبان و مهارت‌های زبانی پرداخته‌اند ملاک نیست. آنچه که مهم است تشخیص میزان توانائی وی در استفاده از زبان بهنگام امتحان می‌باشد. اینکه چگونه به آن درجه از توانائی رسیده است چندان حائز اهمیت نیست.

تست‌های تخصص زبان، در دنیا کاربرد وسیعی دارد. بسیاری از دانشگاهها از آزمون‌های تخصصی بمنظور پذیرش استفاده میکنند. یکی از نمونه‌های خوب آزمون‌های تخصصی زبان تست انگلیسی بعنوان زبان خارجی با نام اختصاری (TOEFL) می‌باشد که محبوبیت جهانی دارد برای تشخیص توانمندی متقاضیان تحصیلات عالیه در کشورهای انگلیسی زبان.

علیرغم استفاده گسترده از آن، ترکیب تست‌های تخصص زبان به نسبت تهیه و تنظیم انواع دیگر آزمون‌ها مشکلتر است. یک دلیل اساسی دیگر این مشکل بودن مربوط میشود به پیچیدگی واژه «تخصصی»^۱ واژه‌ای که دارای تعریف ساده‌ای نیست. استادان معانی مختلفی برای تخصص ذکر کرده‌اند. برای مثال بریه^۲ (۱۹۷۲) چنین تعریفی از آن بدست میدهد «توانمندی نشان دادن مهارت‌ها و معلومات زبان در زمانی واحد بدون برخورداری از کتابی خاص یا فصلی از آن یا شیوه‌های آموزشی». (صفحه ۳۳۲)

چنین تعاریفی موضوع را پیچیده‌تر میسازد، زیرا که تمامی کلمات و عبارات بکار رفته مثل توانمندی نشان دادن، در زمانی واحد و معلومات زبان به خودی خود دارای ابهام هستند. زیرا واژه توانمندی به انواع توانمندی زبانی، اجتماعی و سایر موارد در ارتباط با جنبه تخصصی زبان دلالت میکند (فرهادی)^۳. واژه توانمندی میتواند به توانایی آزمون دهنده در شناخت، درک یا بیان عناصر زبان دلالت نماید. علاوه، در یک زمان واحد، زبان آموز ممکن است علاوه بر شنونده، گوینده یا هر دو باشد. و بالاخره توانمندی نشان دادن میتواند اشکال مختلف داشته باشد، مثل زبان گفتاری، نوشتاری و یا هر دو.

کلارک (۱۹۷۲)^۴ با اصلاح با تعدیلی که در این تعریف بدست داده معتقد است که: «استفاده از زبان با هدف رفع نیاز واقعی زندگی بدون توجه به شیوه ارائه این توانمندی. بدین ترتیب، در آزمون تخصصی (Proficiency) چهارچوب دلالت ... از کلاس به شرایط واقعی در زبان و زندگی بدل میشود». در این تعریف نیز به روشی که زبان آموزی در آن بعمل آمده یکسره نادیده گرفته شده است. بهر حال، به دست دادن هرگونه تعریفی بدون آن که تمامی پارامترهای مربوطه به حساب آورده شود بنظر نوعی ساده انگاری است. این فرض که زبان تخصصی مستقل از روش آموختن آن است محل سؤال است. فرهادی (۱۹۸۳) نشان داد که زبان تخصصی (Proficiency) در واقع مستقل از برنامه‌های آموزشی که ضمن آن زبان آموزان توانمندیهای زبانی را کسب میکنند نیست. او نتیجه میگیرد که انجام تست‌های تخصصی با پس‌زمینه آموزشی، رشته اصلی تحصیلی، جنسیت و ملیت دانش‌آموزان ارتباط نزدیک زیادی دارد. این بمعنی آنست که واژه «تخصصی» بدانگونه که دست‌اندرکاران از آن تعریف به دست میدهند هنوز چندان جامعیت ندارد. همین امر این نکته را که تنظیم تست تخصصی زبان تا چه حد مشکلتر از انواع دیگر تست‌هاست، توجیه مینماید. بهر حال بخاطر اهداف این فصل، همینطور

^۱ - Proficiency

^۲ - Briere

^۳ - Farhady

^۴ - Clark

کافی است تا با اکثر استادان هم نظر شویم که معتقدند که تست تخصصی زبان باید برپایه اندازه‌گیری توانایی آزمون شونده در استفاده از مهارتی ویژه و بمنظور تشخیص میزان عملکرد زبان وی در شرایط واقعی تنظیم یابد.

۳-۲- تست‌های علوم

تست‌های علوم در شرایطی انجام میشود که مدیوم آموزش زبانی غیر از زبان مادری دانشجوست. در این شرایط، زبان دوم زبان آزمون‌هائی است که قرار است میزان معلومات آزمون‌دهنده را در حوزه‌های غیر از زبان بیازماید. مثلاً، تست فیزیک یا روانشناسی تنظیم شده بزبان انگلیسی از جمله تست‌های علمی تلقی میشود. آزمون‌های علمی، نزد آزمون شوندگان زبان از اهمیت نخست برخوردار نیستند. زیرا که این نوع تست‌ها با هدف سنجش توانمندیهای زبانی برنامه‌ریزی نشده‌اند. بلکه بخاطر ارزیابی دانش فرد در یک زمینه علمی و به زبان دوم طرح و تنظیم میشوند. بنابراین، در کشورهائی که علوم بزبان خارجی تدریس نمی‌شود، این گونه آزمون‌ها نقش بخصوصی بازی نمی‌کنند. مثلاً در سیستم آموزشی ایران، آزمون‌های علمی تنها در بخش‌های زبان انگلیسی انجام میگردد و هدف از آنها اندازه‌گیری بنیه علمی دانشجویان در همان حوزه درسی مربوطه می‌باشد. در سایر مراکز تست‌های علمی به همان زبان مادری دانش‌آموزان تهیه و تنظیم میشود.

تمرینات

- ۱- تفاوت بین آزمون تحقیقی و تخصصی در چیست؟
- ۲- تست زبان شامل ساختار، واژگان، درک مطلب شفاهی، درک مطلب خواندنی و انشاء نوشتن می‌باشد. با امتحان گرفتن با یک چنین تستی تشخیص اینکه این تست تحقیقی است یا رتبه‌بندی و یا یراقیشنی چگونه ممکن است؟
- ۳- به مطالبی فکر کنید که مناسب تست استعداد زبان فارسی باشد.
- ۴- تعریف زبان تخصصی را ملاحظه کنید. آیا با تعریف بریه موافقید؟ چرا بله و چرا خیر؟ اگر موافق نیستید، چه عواملی را بر این زبان نزد زبان‌آموزان مؤثر میدانید؟
- ۵- راه‌حل مبنی بر اجتناب از تست‌های رقابتی در افزایش سطح نمرات قبولی یا مردودی است. آیا فکر میکنید که این راه‌حل کمک کند تا مدیران از تست‌های گزینشی استفاده کنند؟ چه راه‌حل‌های دیگری ممکن است ارائه شود؟

اشکال آزمون زبان

۳-۱ مقدمه

آزمون‌دهندگان از همان لحظات اول به‌دست گرفتن دفترچه آزمون متوجه ظاهر آن می‌گردند. اثر این توجه در نگاه نخست بروی کاملاً مهم است. این اثر ممکن است برانگیزنده یا نومیدکننده باشد. او ممکن است ظاهر دفترچه را موافق یا مخالف تصورات خود از آن ببیند.

مثلاً توجه کنید که آزمون شونده در برابر یک سری سئوالات درست یا غلط متن مجبور به پاسخ باشد. تصویری که او پیشاپیش از یک چنین آزمونی می‌داشت حتماً می‌توانست هر چیزی باشد جز یک سری سئوالات درست یا غلط. حتی اگر فرض بر این می‌بود که این نوع سئوالات می‌توانست به سنجش مهارت نوشتاری وی کمک کند، بازهم چنین آزمونی در این وضعیت آزمون‌دهنده را در شرایطی غیرمنتظره قرار میداد. بهمین لحاظ در تهیه و تنظیم یک آزمون معلمان و مربیان می‌باید به شکل آزمون بعنوان یک عامل مهم اهمیت بدهند.

شکل یک آزمون بر چگونگی جلوه ظاهری آن دلالت می‌کند. با توجه به ماهیت و تنوع مهارت‌های زبان، آزمون‌سازان بایست از مناسب‌ترین شکل و متناسب با کیفیت آن چیزی که می‌خواهند بیازمایند در آزمون‌شان سود برند. برای مثال، آزمونی در شکل نوشتاری هرگز مناسب سنجش توانائی درک مطلب شنیداری نیست. و همینطور مناسب ارزیابی معلومات واژگان شفاهی. بنابراین بهنگام تصمیم پیرامون شکل یک آزمون، پارامترهای مشخصی نظیر ماهیت مهارتی که قرار است آزموده شود و چگونگی کارکرد آزمون می‌باید به حساب آورده شوند.

یک آزمون، البته عبارت است از پاره‌ای آیت‌های مشخص و چگونگی شکل این آیت‌های متشکله آن که تعیین کننده شکل آزمون می‌باشد. بدین ترتیب، برای ارائه نگرشی جامع پیرامون قالب آزمون، توضیح ساختار یک آیت بعنوان سنگ بنای آزمون بنظر ضروری مینماید. پس از شرح مفهوم آیت طبقه‌بندیهای متفاوت اشکال آیت ارائه خواهد شد و محاسن و معایب آنها مورد اشاره قرار خواهد گرفت. و سرانجام طبقه‌بندی شکل آیت مطرح خواهد شد.

۲-۳ ساختار یک آیت

یک آیت کوچکترین واحد یک آزمون، شامل دو قسمت است. پایه و پاسخ. منظور از پایه گرفتن اطلاعات از آزمون شونده می‌باشد. پایه میتواند یک سؤال، یک بیان خبری یا سایر ساختارهای زبانی باشد. موارد زیر نمونه‌هایی از پایه‌های مختلف می‌باشد.

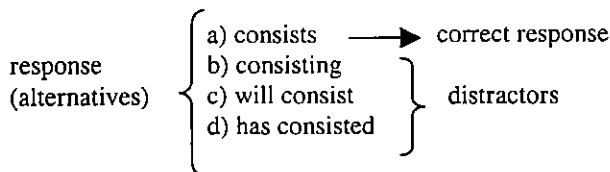
- 1- Explain the purpose of language tests.
- 2- What is the purpose of a placement test?
- 3- Discuss advantages and disadvantages of proficiency tests.
- 4- An aptitude test refers to

هدف از تمامی پایه‌های فوق‌الذکر وادار ساختن آزمون شونده به دادن اطلاعات به امتحان گیرنده می‌باشد.

بخش دوم هر آیت می‌باشد که همان اطلاعات استخراج شده از آزمون شونده می‌باشد. پاسخ متغیری است که ممکن است یک کلمه تنها باشد یا ممکن است یک مقاله جامع مبنی بر بحث پیرامون یک مسئله بفرنج باشد. این آیت بر چگونگی شکل مهارتی که قرار است سنجیده شود، و همینطور بر اهداف یادگیری که آزمون‌دهنده بدان‌ها روزی دست خواهد یافت، و همینطور بر آزمون دیگر زبان که می‌بایست از طریق اشکال درست پاسخ مناسب سؤالات را دریافت کند، متکی است. مثلاً بعضی وقتها از آزمون‌دهنده خواسته میشود تا چند خط یا پاراگراف مطلب بنویسد. در مواقع دیگر، چند پاسخ در اختیار گذاشته میشود تا وی از میان آنها پاسخ درست را برگزیند. مورد نخست که شامل جملات و عبارات پایه می‌باشد و آزمون شونده را وادار به پاسخ درست میکند، در حالیکه مورد دوم نیازمند جملات پایه است و توانایی درک وی را میسنجد.

تمرین گونه دوم در شکل تمرینات چند گزینه‌ایها متداول است. در آیت‌های چند گزینه‌ای، یعنی متداولترین شکل تمرین، بدنبال بُن تمرین سه، چهار یا پنج پاسخ آورده میشود. این پاسخ‌ها ضمناً انتخاب، گزینش یا همان گزینه خوانده میشوند. که یکی از آن میان پاسخ است، و بقیه انحرافی هستند. نمونه زیر اجزای یک آیت چند گزینه‌ای را نشان میدهد.

An item. of two parts.



نمونه فوق تفاوت بین یک گزینه درست و گزینه‌های انحرافی را روشن می‌سازد. یک گزینه ممکن است درست یا نادرست باشد. عبارت دیگر گزینه هم می‌تواند پاسخ درست باشد و یا انحرافی، در حالیکه انحرافی‌ها تنها گزینه‌های غلطند.

اینک، ارائه چگونگی ساختار یک آیتم، خواننده باید بخاطر داشته باشد که شکل یک آزمون توسط شکل آیتم‌های تشکیل دهنده آن مشخص می‌شود. بنابراین، هر طبقه‌بندی‌ای که بلحاظ شکل آیتم بدست دهیم در عین حال می‌تواند طبقه‌بندی شکل آزمون هم به حساب آید.

۳-۳ طبقه‌بندی اشکال آیتم

تا به حال در مورد شکل آزمون در این زمینه سردرگمی‌هایی وجود داشته است منشاء این سردرگمی یکی استفاده از واژگان مختلف برای دلالت بر یک مفهوم یکسان یا مشابه می‌باشد. برای مثال، واژگانی نظیر fill in the blanks supply و Completion تا حال بمنظور دلالت بر شکل واحدی از آیتم مورد استفاده بوده است. منشاء دیگر این سردرگمی ناشی از نظر گاه‌های متفاوت استادان در استنباط از مفاهیم می‌باشد. مثلاً، طبقه‌بندی آزمون‌ها به اشکال عینی و ذهنی ناشی از تفاوت مکتب‌های فکری در مورد چگونگی آیتم‌هاست. علیرغم مشابهت‌های موجود در طبقه‌بندی، هر یک دارای محاسن و معایبی است. بنابراین، با ارائه طبقه‌بندی‌های متفاوت و بررسی انتقادی آنها پیشنهاد مبنی بر طبقه‌بندی جدید در فصل بعد به روشن ساختن مسئله خواهیم پرداخت.

۳-۳-۱ آیتم‌های ذهنی و عینی

از آغاز دوره که روش ترجمه - گرامر آموزش زبان پایه گذاشته شد تاکنون آزمون‌های ترجمه زبان بعنوان عمده‌ترین تکنیک امتحانی مورد استفاده قرار گرفت. این آزمون‌ها از آزمون شونده خواستار آند تا متنی را ترجمه یا یک سری جملات را از زبانی به زبان دیگر برگردانند. گرچه محتوای چنین آزمون‌هایی بنظر به مواد درسی که قرار بود امتحان شود مربوط بود و در بیشتر موارد بواقع نمونه مطالب درسی بود، روش نمره دادن به این تست‌ها سیستمیک نبود. نمره دادن به تست‌های ترجمه مستلزم صرف مقدار زیادی زمان و نیز انرژی است. بعلاوه، نوسان نمرات از ممتحنی به ممتحن دیگر بر

سرراه یکنواختی نمرات همیشه یک مشکل است. از آنجائیکه روش‌های نمره دادن به چنین تست‌هایی از معیار عینی‌ای تبعیت نمیکنند، بهمین‌خاطر جزء آزمون‌های نظری یا سلیقه‌ای به حساب می‌آیند. بمنظور جبران نقایص آزمون‌های ذهنی از طرفی و بهره‌جستن از اصول روان‌سنجی^۱ در آزمون زبان از سوی دیگر، استادان به سمت آزمون‌هایی با نمرات یکنواخت و ساده‌گرایش پیدا کردند. بخاطر این وجه تفاوت با آزمون‌های تستی ذهنی^۲، این تست‌های بتازگی طرح و تنظیم شده در طبقه‌بندی بنام آزمون‌های عینی یا ایژکتو^۳ خوانده شده‌اند. چند گزینه‌ای‌ها و انواع درست - غلط از نمونه‌های مسلم معیارهای عینی در آزمون‌سازی می‌باشند.

باید اذعان کنیم که تفاوت بین تست‌های عینی و ذهنی، که ریشه در تفاوت بین آزمون‌های ترجمه و تست‌های چند گزینه‌ای دارد، در این زمینه منجر به سوءتفاهمی رایج گردیده است. این سوءتفاهم از آنجا ناشی میشود که گمان میرود بحث ذهنیت و عینیت مربوط به روش‌های نمره دادن است و نه شکل یک آزمون. مثلاً این از همان درک نادرست سرچشمه میگیرد که فکر کنیم تمام آزمون‌های انشاءای ذهنی‌ند و تمامی چند گزینه‌ایها تست‌های عینی هستند. مثال‌های زیر بروشنی گویای این واقعیت است که شکل یک آزمون هیچ ربطی به ذهنی بودن یا عینی بودن آن ندارد.

مثال: (۱)

There are four in a year.

- a) seasons b) months c) weeks d) fortnights

Example 2

There are four seasons in a year.

True false

Example 3

How many seasons are there in a year?

Answer: 4

Example 4

In summer, The weather is usually hot here.

True false

تمامی این آیتم‌ها، بدون توجه به شکل آنها، عینی هستند، زیرا که به تمامی دارای پاسخ‌های ثابت و صحیح و لامتغیرند و نمرات بندرت تابع سلاقی و نظریات ممتحنین هستند. با این حال، آیتم‌های زیر، که بلحاظ شکل با آیتم‌های فوق شباهت دارند عینی نیستند.

Example 5

There are four Seasons in a year.

- a. beautiful b. tiring c. uninteresting d. exciting

^۱ - Psychometric

^۲ - Subjective

^۳ - objective

Example 6

There are four beautiful seasons in a year.

True false

Example 7

How many exciting seasons are there in a year?

Example 8

In summer, the weather is usually pleasant here,

Example 9

Which season do you like best?

دو سری آیتم بلحاظ شکل خیلی مشابه هستند. در عین حالی که در مجموعه اولی (مثال‌های ۱ تا ۶) عینی هستند، در مقابل مجموعه دوم (مثال‌های ۷ تا ۱۱) کاملاً ذهنی می‌باشند. بعبارت دیگر، پاسخهای دسته اول آنچنانند که تقریباً هر کسی یک پاسخ مشخص بدان خواهد داد. در حالیکه پاسخهای آیتم‌های دوم به میزان زیادی مبتنی است بر سلاقی، نظریات، دوست داشتن و نداشتن‌های افراد ممتحن. بدین‌ترتیب، تقسیم‌بندی آیتم‌ها به عینی و ذهنی مسئله‌ای نیست که به شکل آیتم مربوط باشد، بلکه موضوع عینی بودن یا ذهنی بودن ارتباط به روش نمره دادن دارد. از میان دو نمونه مثال با شکل یکسان، یکی کاملاً عینی و دیگری کاملاً ذهنی است. بنابراین، یک طبقه‌بندی متفاوت ضرورت می‌یابد تا بموارد ابهام ناشی از دوگانگی در مقوله عینی، ذهنی خاتمه داده شود.

۳-۳-۲ نارسائی تست تشریحی در مقابل آیتم‌های چندگزینه‌ای

از آنجائیکه نارسائی تقسیم‌بندی آیتم‌ها به عینی و ذهنی ثابت ساخت که تمامی اشکال آیتم‌ها را نمیتواند شامل شود، استادان در صدد تحول نگرش جدیدی در این زمینه برآمدند. آنها آیتم‌ها را بدو طبقه تقسیم نموده‌اند: تشریحی و چندگزینه‌ای. آیتم‌های توضیحی - تشریحی به تمامی انواع آیتم‌هایی گفته میشود که از آزمون شونده خواسته میشود تا عناصر زبانی همان آیتم‌های چندگزینه‌ای را خود بنویسد. همینطور به تمامی آیتم‌هایی دلالت میکند که آزمون شونده باید از میان گزینه‌های مختلف پاسخ درست را برگزیند. این طبقه‌بندی نسبت به طبقه‌بندی قبلی یک پیشرفت محسوب می‌شود، با این همه دارای نواقص چند می‌باشد. به مثال‌های زیر توجه کنید:

1. The Students play now.
2. The students in the park everyday.
3. What is the past tense of "go" ?
4. what is the passive form of "He wrote a book" ?
5. Change the following sentence into indirect form:
Hossein said , "I'm going to start my graduate studies in physics at teheran university."
6. Write a short paragraph about "your education".
7. Discuss the advantages and disadvantages of living in big cities.
8. Compare and contrast two educational systems in Iran.

تمامی آیتم‌های فوق‌الذکر، از نوشتن یک واژه تنها گرفته تا تشریح کامل یک مفهوم یا موضوع، انواع سؤالات تشریحی محسوب میشوند. از آنجائیکه هر یک از این آیتمها نیازمند نوع معینی از فعالیت می‌باشند، نمیتوانند تحت یک طبقه‌بندی یکسان نام برده شوند. پس بدیهی است، طبقه‌بندی آزمون تحت‌عنوان سؤالات تشریحی بنظر چندان موافق نمی‌آید، زیرا که توضیح یا تشریح شکل خاص خودش را داراست. نوشتن یک واژه تنها بطور منطقی نمیتواند با نوشتن یک توضیح یا مطلب یکسان تلقی گردد. از این‌رو، طبقه‌بندی اشکال آیتمها نیازمند اصلاح یا تعدیل بیشتری است.

۳-۳-۳ آیتم‌های کامل کردنی در مقابل آیتم‌های تشخیصی

برای اجتناب از نواقص طبقه‌بندی قبلی، آیتم‌های جدیدی مطرح گشته بدون آنکه تغییرات قابل توجهی در اصل طبقه‌بندی رخ داده باشد. عبارات «عینی» یا «چند گزینه‌ای» جای خود را با «تشخیص دادن» عوض کرده‌اند. آیتم‌های تشخیصی تنها نیازمند است تا آزمون شونده شکل درست را از میان گزینه‌های مختلف بشناسد و انتخاب کند. از طرف دیگر واژگانی نظیر Suppletion و گاهی Completion تنها از آزمون‌دهنده می‌خواهد تا قسمت خالی سؤال را پر کند یا جمله ناقص را کامل نماید.

با وجود آنکه این طبقه‌بندی طبقه‌بندی قبلی را تعدیل نموده است، اما همچنان به نواقص مشابه دچار است به مثال‌های زیر توجه کنید:

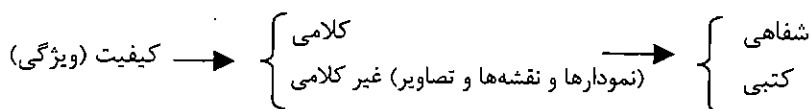
- 1- Hossein will His education abroad. (continue)
- 2- Parviz will
- 3- Write a paragraph about "the process of reading"

تمامی موارد بالا از نوع بهترین‌های تکمیلی یا کامل کردنی هستند. با این وجود حدود تکمیل کردن و روش نمره دادن به آنها روشن نیست. باستثنای مورد اول، که پاسخ به آن معلوم می‌باشد، بقیه ایراداتی مشابه آیتم‌های تشریحی - توضیحی دارند. بدین‌ترتیب ادامه این طبقه‌بندی نمیتواند معرف تمامی موارد متفاوت باشد. برای رفع پاره‌ای ابهامات در رابطه با تقسیم‌بندی آیتمها، و پرتوافکندن بر مسئله، یک طبقه‌بندی تازه بگونه زیر پیشنهاد شده است. امید میرود تا نگرش جدید بدلائل نظری و عملی جامع باشد.

۴-۳ طبقه‌بندی زبان - روانی

باید اشاره کرد که شکل یک آیتم بدون بررسی انتقادی چگونگی آن نمیتواند معین گردد. یک آیتمی را چند گزینه‌ای، کامل شونده یا هر نامی که میخواهید برآن بگذارید این امر بدون تشخیص فرایندی که به آیتم پاسخی گفته میشود به طبقه‌بندی منطقی‌ای منجر نخواهد شد. بنابراین در طبقه‌بندی پیشنهادی زیر شکل آیتم از طریق توجه به قواعد نظری فرایند زبان قابل تبیین است.

طبقه‌بندی حاضر بدلیل آنکه اصول روانشناسی و زبانشناسی را بعنوان مبانی نظری و انسانی شکل آیتم فرض میکند، روان زبانی خوانده میشود. این طبقه‌بندی از اصول روانشناسی بهره میجوید زیرا که پاسخ به یک آیتم نیازمند فرایندهای روانشناختی معینی است. علاوه بر این، طبقه‌بندی پیشنهادی مستلزم نظریات زبان شناسی است زیرا آیتمی که برای ارائه و کمک به یکی از اشکال معین زبان پاسخ داده میشود نیازمند استفاده از عناصر زبانشناسی است. با توجه به روانشناسی، از همان نخستین لحظه مواجه با یک آیتم فرایندهای روانشناختی نظیر ادراک، تشخیص و شناسائی، درک، تحلیل و ارائه مطالب نوشتاری یا شفاهی بطریقی مورد استفاده قرار میگیرند. بنابراین هر معیاری برای تشخیص شکل آیتم بایست فرایندهای روان شناختی مورد لزوم برای پاسخگویی به آن آیتم را مدنظر داشته باشد. با توجه به اصول زبانشناسی، زبان بدو طریق ممکن است گل کند که ما به این دو طریق، کیفی و ابزاری اطلاق می‌کنیم. کیفیت زبان مربوط میشود به گونه گونی تجلی آن نظیر شفاهی، نوشتاری و اشکال تصویری یا تجسمی. به سخنی دیگر، زبان میتواند بصورت کلامی و غیرکلامی متجلی گردد. شکل کلامی زبان شامل فرم شفاهی و نوشتاری و غیرکلامی عبارت است از انواع شیوه‌های گرافیکی. نمودار زیر روشنتر این امر است.



بدین ترتیب، معیار دیگر برای تشخیص شکل آیتم همان چگونگی شکل ارائه آن می‌باشد. پس بدین ترتیب طبقه‌بندی روان زبانی دو وجهی خواهدبود. یک وجه ناظر بر فرایند روانشناختی که در امر پاسخگویی به یک آیتم دخالت دارد و وجه دیگر که به کیفیت زبانی ارائه آیتم مربوط است. برای اجتناب از پیچیدگی موضوع، دو فرایند روانشناختی، که در فرایند زبان اهمیت دارند، مثل شناختن و ارائه دادن بعنوان دو قطب ابتدا و انتهای عملکرد و فرایندی زبان تلقی میشوند، و از دیگر فرایندهای روانشناختی یعنی درک کردن بین آن دو جای دارد.

البته، شناختن پیش‌شرط درک کردن و درک کردن پیش‌شرط بیان کردن است. یعنی اینکه، فرایند زبان از شناخت شروع میشود و تا ابداع ادامه میابد. از آنجا که در این میان تصور فضائی خالی ممکن نیست، پس تداوم پابرجاست با عمل شناختن‌ها در ابتدا بیان کردن‌ها در انتها. علاوه بر این بدلیل ممکنات بین این دو مراحل درک کردن و درک و بیان نباید از نظر دور داشت. این فرایند، فعالیت‌هایی را دربرمیگیرد که مستلزم میزانی از درک و بیان است.

پس وجه روانشناختی شامل مواردی میشود که نیازمند شناختن، درک و درک و بیان یا بیان تنها از جانب آزمون‌دهنده باشد. وجه زبان شناختی، نیز شامل وجوه گفتاری، نوشتاری و شکل تصویری یا تجسمی زبان می‌باشد.

علاوه بر فاکتورهای زبان شناختی و روانشناختی، یک تیغه دیگر، یعنی اجزاء یک آیتم، باید به حساب بیاید. همانطور که قبلاً اشاره شد، یک آیتم شامل اصل و پاسخ است. این اجزاء به عامل دخیل در تشخیص شکل آیتم وضعیتی متغیر میدهد. با توجه به همین امر اصل و پاسخ هم ممکن است و هم ممکن نیست که دارای حالت زبانی یکسانی باشند. بعبارت دیگر، حالت پایه ممکن است از حالت پاسخ فرق کند که در اکثر موارد فرق میکند. بنابراین آیتم‌ها برحسب فرایند روانشناختی و حالت زبانشناختی پایه و هم پاسخ ممکن است دارای اشکال متنوعی باشند.

این فاکتورها، اعم از زبانشناختی و روانشناختی بعلاوه ترکیب آیتم در جدول زیر نشان داده شده‌اند.

جدول ۱-۳

زبان‌شناسی روانشناسی	توانایی شناخت	توانمندی درک	درک / نوشتن	نوشتن (تولید)
شفاهی	۱	۲	۳	۴
کتبی	۵	۶	۷	۸
تصویری	۹	۱۰	۱۱	۱۲

برای جدول ۱۲ خانه تنظیم شده است: هر یک بمعنای یک وجه آیتم می‌باشد. مثلاً، یک آیتم درست - غلط بمنظور آنکه هدف از طرح آن برآورده شود به شکل زیر ارائه میشود.

a) فرایند روانشناختی: شناختن یا درک آن

b) پایه: گفتاری، نوشتاری، یا تصویری - تجسمی

c) پاسخ: گفتاری، نوشتاری یا تصویری - تجسمی

بنابراین، یک چنین آیتمی در صورتی که قرار باشد توانائی شناخت از آزمون دهنده را بسنجد در یکی از خانه‌های ستون ۱ و اگر بخواهد توانمندی درک وی را بیازماید در یکی از خانه‌های ستون ۲ جای خواهد گرفت. بعلاوه، در صورتی که به شکل گفتاری باشد در خانه‌هایی ۱ یا ۲ و در شکل نوشتاری در خانه‌های ۵ یا ۶، و چنانچه شکل تصویری - تجسمی داشته باشد در خانه‌های ۹ یا ۱۰ جای میگیرند.

تمامی خصوصیات آیتم بر حساب جای آن در یکی از خانه‌های جدول میتواند معین گردد. مثالی دیگر، آزمونی مثل دیکته نیازمند هم درک و هم بازنویسی از جانب آزمون‌دهنده می‌باشد. اینکه آیا پایه و پاسخ به شکل گفتاری یا نوشتاری است بر تنوع شکل آزمون میافزاید. چنین آزمونی در یکی از خانه‌های ستون ۳ برحسب حالت آن قرار خواهد گرفت. و بالاخره، آزمونی همچون انشاء اختیاری نیازمند تنها نوشتن می‌باشد که در این صورت در یکی از خانه‌های ستون ۴ جدول جای خواهد گرفت.

با توجه به فاکتورهای فوق‌الذکر، آزمون سازان بایست در نامگذاری یک آیتم بنام چند گزینه‌ای یا توضیحی، تشریحی احتیاط زیادی بخرج دهند. بنظر خیلی ساده انگاری است اگر تصور کنیم که یک

آیتم را می‌شود به این شکل یا آن شکل، بدون در نظر گرفتن متغیرهای فوق‌الذکر طبقه‌بندی کرد. بنابراین قبل از اینکه شکل هر آیتمی را نامگذاری کنیم دقت بیشتری باید بخرج دهیم.

فعالیت‌ها:

تمرین

- ۱- تفاوت بین گزینه درست و گزینه انحرافی چیست؟
- ۲- به چه ترتیب شکل یک آزمون بر محتوا یا عملکرد آن تأثیر می‌گذارد؟
- ۳- به چه طریق نمره دادن به تست‌های تشریحی بطور عینی امکان دارد؟
- ۴- امتیازهای طبقه‌بندی روان‌بانی بر دیگر طبقه‌بندیها در چیست؟
- ۵- جدول مربوط به طبقه‌بندی روان‌بانی را ملاحظه کنید، شکل‌های آیتمی را از نظر بگذرانید که میتواند در هر خانه جدول جای گیرد.

فصل چهارم

روش‌های اساسی آمار در آزمون

۱-۴ مقدمه

در سال‌های اخیر دانشکده‌های هر چه بیشتری به ارائه دوره‌های آمار مقدماتی مبادرت ورزیده‌اند، هدف از طرح این مبحث در این فصل ارائه درآمدی عمومی بر آمار موردنیاز دانش‌جویان در زمینه آموزش زبان بطور عام، و از آزمون آن بطور خاص می‌باشد. بهر صورت دانش‌جویان می‌بایست معلومات خود را در زمینه زبان و شیوه‌های آماری ارائه دهند. بر همین اساس تمرینات و مثالهایی برای جلب توجه آنان برگزیده شده است.

ما اکثراً بدین گمان هستیم که علم آمار با نقشه و جداول ارقام و اعداد سروکار دارد. در عین حالیکه این گمان غلطی هم نیست. کارشناسان ریاضی و آمار واژه آمار، را در معنای عمومی آن بکار می‌برند. بطور کلی، واژه آمار بدانگونه که مورد استفاده کارشناسان این رشته قرار می‌گیرد مستلزم جمع‌آوری اطلاعات به اتکاء ارقام یا داده‌ها و همینطور تحلیل و بررسی آنها، و اتخاذ تصمیم براساس نتایج بررسی‌ها می‌باشد.

نقشی که آمار در زندگی و فعالیت‌های روزانه ما بازی میکند پیوسته رو به افزایش است. به واقعیت تاریخ برگردیم. دانشمند قرن نوزده اچ جی ولز^۱ روزی پیش‌بینی میکرد که «تفکر آماری مثل توانایی خواندن و نوشتن لازمه کارآمدی زندگی شهروندی است.» مثال‌های زیر درباره استفاده از اطلاعات آماری امروزه به یک ابزار مهم برای هر شخص تحصیل کرده‌ای درآمده است.

^۱ - H. G. Wells

- ۱- آخرین آمار بانک مرکزی نشان می‌دهد که هزینه زندگی در توکیو ۸ درصد در سال افزایش دارد.
 - ۲- آمار حاکی از آن است که استفاده از بعضی از داروها می‌تواند برای نوزادان مخاطره‌آمیز باشد.
 - ۳- آمار جمع‌آوری شده توسط شرکت‌های بیمه نشان از آن دارد که مردمان امروز میانگین عمری بالاتر از پدر و مادرهایشان دارند.
 - ۴- آمارها حاکی از آنست که هوایی که ما تنفس می‌کنیم بشدت آلوده می‌باشد، از این رو بی‌هیچ تردیدی پاره‌ای از مردم دچار بیماری‌های ناشی از هوای آلوده‌اند.
 - ۵- آمارها نشان می‌دهند که سرعت جمعیت جهان از مکانیزم تهیه غذا پیشی گرفته است.
 - ۶- آمار حاکی از آن است که هر دانش‌جوئی، بدون توجه به سابقه تحصیلات دبیرستانی، چنانچه بخوبی مشوقشان باشند در تحصیل در دانشگاه هم موفقند.
- برای جمع‌آوری اطلاعات و تنظیم جداول داده‌ها به معلومات ریاضی معمولی و خیلی کمی احتیاج است. با این همه، تحلیل اطلاعات بشیوه‌ای معنادار محتاج بررسی دقیق است. در صورتی که داده‌ها توسط کارشناسان ریاضی و آماری مجرب مطالعه و بررسی نگردند، آمار ارائه شده گمراه کننده خواهد بود.
- در شهر (*) حادثه پولیو که تصور میشد واقعیت ندارد از سال ۱۳۶۴ تا ۱۳۶۵ صددرصد افزایش یافت. یک چنین آماری والدین را به وحشت می‌انداخت. بهر صورت، برحسب بررسی‌های دقیق تازه معلوم شد در سال ۱۳۶۴ در میان جمعیت ۱۰۰۰/۰۰۰ نفری تنها دو مورد و در سال ۱۳۶۵ ۴ مورد گزارش شده است.

۲-۴ اقدام به عمل برحسب توصیه‌های آماری

واژه آمار موارد استفاده مختلفی دارد. کارشناسان ریاضی آمار را بدو حوزه آمار تشریحی، و آمار استنباطی تقسیم‌بندی میکنند.

روش آماری که از داده‌ها به اختصار نمونه برمیدارد، آمار توصیفی خوانده میشود (هاج و فرهادی ۱۹۸۲). در این روش ما به توصیف نمونه آماری می‌پردازیم. هر یک از مشخصات این نمونه را آمار مینامیم. و میتوانیم از طریق این آمار یعنی وجوه مشخصه یک نمونه درباره ویژگیهای آماری یک جمعیت معین که پارامتر گفته میشود اظهارنظر کنیم. این امر بطریق استفاده از شیوه‌های آمار استنباطی ممکن میگردد.

پیچیدگی آمار استنباطی خیلی بیشتر از آمار توصیفی است. برای درک آنچه که در این امر دخالت دارد، تصور کنید که علاقمند باشیم که معدل توانائی زبان گفتاری دانش‌جویان یک دانشگاه را بدانیم. از آنجا که تعداد دانش‌جویان حضوری یک دانشگاه بسیار زیاد است، و مصاحبه با هر دانش‌جوئی، در نتیجه جمع‌آوری اطلاعات لازم در این زمینه کار بس گسترده‌ای است. علاوه بر این، هزینه این کار بدون تردید بالا و در عین حال بسیار وقت‌گیر هم هست، ما تا حد ممکن میتوانیم اطلاعات لازم را از یک

نمونه آماری در حدی که کفایت کند و پاسخ مناسبی به نیاز ما در این زمینه باشد کسب کنیم. پس میتوانیم بمنظور پیش‌بینی پیرامون تمامی پیکره دانشجویی و به سخنی جمعیت دانشجویان اظهارنظر نمائیم. این همان چیزی است که استنباط آماری خوانده میشود.

۳-۴ جدول‌بندی داده‌ها

قبل از آنکه داده، بلحاظ کمی مورد مطالعه قرار گیرد، معمولاً لازم می‌آید تا خلاصه‌ای از آن بدست داده شود. جدول ۴-۱ نشان از رکورد کلاسی مبنی بر تست قرائت که در آغاز سال تحصیلی برگزار شده دارد. نمرات بهمان ترتیب که در دفتر نمرات کلاسی معلم درج گردیده در جدول آمده است. با این وجود نمرات به این شکل چندان معنایی ندارند، و ما تنها با قدری اشکال می‌توانیم بگوئیم که مثلاً آیا دانش‌آموز ردیف نخست لیست، با نمره ۹۰ در مقابل نمره ۱۲۸ در زمینه آمادگی قرائت به نسبت همکلاسیها برتری دارد یا دانش‌آموز متوسطی است. جدول ۴-۱

Student	Score	Student	Score	Student	Score	Student	Score
1.	90	11.	59	21.	75	31.	81
2.	66	12.	95	22.	75	32.	71
3.	106	13.	78	23.	51	33.	68
4.	84	14.	70	24.	109	34.	112
5.	105	15.	47	25.	89	35.	62
6.	83	16.	95	26.	58	36.	91
7.	104	17.	100	27.	59	37.	93
8.	82	18.	69	28.	72	38.	84
9.	97	19.	44	29.	74		
10.	97	20.	80	30.	75		

جدول ۴-۱

۳-۴-۱ به ترتیب رتبه

عموماً اولین گام در تنظیم لیست نمرات رعایت کمیت آنها بطور معمول از بالاترین به پائین‌ترین سطح می‌باشد. در یک کلاس کوچک، این امر اغلب همان چیزی است که ضرورت دارد. جدول ۴/۲ نشان می‌دهد که ۳۸ نمره بترتیبی کمی از شماره ۱۱۲ تا ۴۴ تنظیم یافته است. این جدول در ضمن بیانگر آنست که دانش‌آموزان رتبه‌بندی (از رتبه اول تا رتبه سی‌هشتم) و نمرات بدون گروه‌بندی بیشتر جدول‌بندی شده‌اند. اکنون بسادگی میشود فهمید که نمره ۹۰ یک دانش‌آموز او را در ردیف ۱۳ یک کلاس ۳۸ نفری قرار میدهد. یا حدوداً در ردیف یک سوم کلاس از بالا. بهمین ترتیب تفسیر هر

یک از نمرات دیگر بلحاظ رتبه، کاری است آسان. اما تساوی نمرات هم به تعیین پیش میاید، بخصوص در کلاس‌های ۲۰ تائی یا بیشتر. مثلاً توجه کنید، که دو دانش‌آموز بطور یکسان نمره ۹۷ گرفته‌اند. از آنجا که با عقل خود جور در نمی‌آید که بگوئیم یکی بالاتر از دیگری است، بنابراین باید رتبه واحد به هر دو تا بدهیم. و چون شش دانش‌آموز هستند که در ردیف بالاتری قرار دارند (۶ و ۵ و ۴ و ۳ و ۲ و ۱)، میانگین دو ردیف بعدی، یعنی ۷ و ۸ میشود $7/5$. بهمین ترتیب میانگین رتبه ۹ و ۱۰، $9/5$ است و بهمین‌ترتیب برای دانش‌آموزانی با نمرات تکراری سه دانش‌آموز با نمرات ۷۵ وجود دارند، و ۲۱ دانش‌آموز که بالاتر از این نمره هستند؛ میانگین سه رتبه دیگر (۲۲ و ۲۳ و ۲۴) ۲۳ می‌باشد، که رتبه‌ای است متعلق به نمرات تکراری ۷۵. علاوه بر صرف وقت و زحمتی که بهنگام تعیین این رتبه‌ها لازم میاید، لیست بلند بالا برای مقایسه با کلاسهای دیگر که یا خیلی بزرگترند یا خیلی کوچکتر هیچ تناسبی ندارد؛ مثلاً رتبه نوزده در یک کلاس ۳۸ دانش‌آموز خیلی ضعیفتر از همان رتبه در یک کلاس ۷۰ نفری است. توزیع فراوانی برای رفع همین مشکلات مطرح شده است.

۲-۳-۴ توزیع فراوانی

فهرست نمرات از طریق تنظیم نمرات بشیوه توزیع فراوانی، یا معمولاً بطور خلاصه بگوئیم توزیع، میتواند کوتاه‌تر گردد. ستون سوم و چهارم جدول نشاندهنده ساده‌ترین شکل توزیع می‌باشد. نمرات مختلفی بترتیب کمیت تنظیم میشوند، در اینجا ۱۱۲ تا ۴۴ و در سمت راست هر یک از نمرات تعداد دفعات تکرار آنها یادداشت شده است. درج هر رقم در سمت راست نمره تعداد فراوانی یا بطور اختصار (f) خوانده میشود. و مجموع فراوانیها با حرف (N) نمایش داده میشود.

Order of Size	Rank	Score	Frequency (f)	
112	1	112	1	
109	2	109	1	
106	3	106	1	
105	4	105	1	
101	5	104	1	
100	6	100	1	
97	7.5	97	2	Sum = 19
97	7.5	95	2	
95	9.5	93	1	
95	9.5	91	1	
93	11	90	1	
91	12	89	1	
90	13	84	2	
89	14	83	1	
84	15.5	82	1	
84	15.5	81	1	
83	17			Midpoint of frequencies
82	18	80	1	
81	19	78	1	
80	20	75	3	
78	21	74	1	
75	23	72	1	
75	23	71	1	
75	23	70	1	
74	25	69	1	Sum = 19
72	26	68	1	
71	27	66	1	
70	28	62	1	
69	29	59	0	
68	30	58	1	
66	31	51	1	
62	32	47	1	
59	33.5	44	1	
59	33.5			
58	35			N=19+19=38
51	36			
47	37			
44	38			

۳-۳-۴ تشخیص درصدی

تشخیص وضعیت توزیعی یک نمره اغلب با اهمیت تلقی می‌شود. مثلاً، تصور کنید که امتیاز دانش‌آموزان همان نمرات عددی در کلاس درس باشد و ما بخواهیم چگونگی وضعیت درسی یک دانش‌آموزی را مشخص سازیم. از گزارش به سادگی برمیایم که «رضا سومین نمره بالا را در کلاس کسب نموده است.» یا امیر بعد از او پائین‌ترین نمره را گرفته است» آنچه که از گزارش برمیاید این است که تعداد دانش‌آموزان در هر سمت هر یک از این نمرات معلوم است.

با این وجود اطلاعات کسب شده در این باره مشخص نمی‌کند که رضا یا امیر تا چه حد موفق بوده‌اند. هر کس می‌تواند فکر کند که تعداد دانش‌آموزان چهار نفر بوده‌اند یا یکصد نفر. برای اینکه معین کنیم که وضعیت درسی رضا یا امیر در گروه کلاسی به چه منوال بوده است کل دانش‌آموزان کلاس باید معلوم گردد. از این‌رو وضعیت یک نمره نبایستی بهمین سادگی که چند نمره بالای آن و یا چند نمره زیر آن قرار دارد اعلام شود. تعداد نمرات در تمامی فراوانی (گروه) بایستی نیز مشخص گردد. ما هم‌اکنون چگونگی وضعیت قرار گرفتن یک نمره را در یک فراوانی بالارزش دادن به آن یا استناد به رتبه درصدی، مورد توجه قرار خواهیم داد. اما قبل از اینکه جلوتر برویم، مفهوم دو عبارت دیگر می‌باید روشن گردد از جمله:

الف) فراوانی نسبی

ب) فراوانی کلی

۱-۳-۳-۴ فراوانی نسبی

فراوانی نسبی اشاره به فراوانی هر نمره تقسیم بر تعداد کل نمرات دارد. تصور کنید در یک کلاس انگلیسی تعداد دانش‌آموزان ۶۰ نفر باشد ($N = 60$). نمرات (از میان یکصد نفر) گزارش شده توسط معلّم همانطور که در جدول ۳-۴ آمده چنین است.

جدول ۳-۴ فراوانی نسبی

نمره	فراوانی	فراوانی نسبی		درصد
		محاسبه	نتیجه	
۸۰	۵	۵ : ۶۰	۰/۰۸	$۱۰۰ = ۸ * ۰/۰۸$
۷۰	۱۵	۱۵ : ۶۰	۰/۲۵	$۱۰۰ = ۲۵ * ۰/۲۵$
۶۰	۲۰	۲۰ : ۶۰	۰/۳۴	$۱۰۰ = ۳۴ * ۰/۳۴$
۵۰	۱۵	۱۵ : ۶۰	۰/۲۵	$۱۰۰ = ۲۵ * ۰/۲۵$
۴۰	۵	۵ : ۶۰	۰/۰۸	$۱۰۰ = ۸ * ۰/۰۸$

از آنجا که ما سروکارمان با دانش‌آموزان می‌باشد، نمی‌توانیم بگوئیم که ۸٪ دانش‌آموزان آزمون را از سرگذرانده‌اند یا چه نمره خاصی گرفته‌اند. بنابراین فراوانی نسبی معمولاً در ۱۰۰ ضرب میشود تا نتیجه درصدی بدست آید. آنگاه گفته میشود که ۸ درصد موضوع تحقیق آزمون را گذرانده‌اند یا نمره مشخص دریافت کرده‌اند.

۲-۳-۳-۴ فراوانی کلی

فراوانی کلی نشاندهنده وضعیت معین هر نمره در گروه نمرات می‌باشد. این فراوانی نشان می‌دهد که چه تعداد نمره در زیر آن در جدول توزیع قرار دارد. این فراوانی همچنین برای محاسبه نمرات درصدی بکار میرود. جدول ۴-۴ نمرات ۱۲۵ نفر از دانش‌آموزان را در یک آزمون انگلیسی ۴۰ آیتی جایگذاری نشان میدهد.

جدول ۴-۴ نمرات آزمون جایگذاری زبان انگلیسی

نمرات آزمون	فراوانی (F)	فراوانی تراکمی (F)	درصد
۳۸	۱	۱۲۵	$(100) \frac{125}{125} = 100$
۳۷	۱	۱۲۴	
۳۶	۳	۱۲۳	
۳۵	۵	۱۲۰	
۳۴	۹	۱۱۵	
۳۳	۸	۱۰۶	
۳۲	۱۷	۹۸	
۳۱	۲۳	۸۱	
۳۰	۲۴	۵۸	
۲۹	۱۸	۳۴	
۲۸	۱۰	۱۶	$(100) \frac{81}{125} = 64/8$ یا ۶۴٪
۲۷	۳	۶	
۲۶	۱	۳	
۲۵	۵	۲	
۲۴	۲	۲	

$$N = 125$$

همانطور که در این جدول نشان داده شده، مقادیر ستون سوم زمانی به دست می‌آید که ما در کار قبلی فراوانی نمره‌های متوالی را بگنجانیم. ستون از انتها به بالا تنظیم شده و در ورودی انتهایی دو نمره وجود دارد و نمره صفر مجاور نمره انتهایی است. بنابراین برای محاسبه فراوانی کلی عدد ۲ به صفر اضافه شده و نظایر آن.

در جدول، فراوانی ساده، یا فراوانی مطلق با حرف (f) و فراوانی کلی با حرف (F) نشان داده می‌شود. بمنظور محاسبه رتبه درصدی در هر سطح یا نقطه‌ای، (F) همطراز بایستی تقسیم بر تعداد کل نمرات یا دانش‌آموزان یعنی (N) گردد. نتیجه در ۱۰۰ ضرب می‌شود. برای مثال، در جدول بالا فراوانی دانش‌آموزانی که نمره ۳۱ دارند ۲۳ می‌باشد. فراوانی کلی عدد ۸۱ می‌باشد. تعداد کل نمرات یا دانش‌آموزان یعنی عدد (N) ۱۲۵ می‌باشد. بنابراین رتبه درصدی (P) برای این سطح چنین است:

$$P = 100 \frac{F}{N} = 100 \frac{81}{125} = 64.8 \text{ or } 65\% \text{ (نمره گرد شده)}$$

بدین ترتیب رتبه درصدی در این مورد بمعنی آنست که ۶۵ درصد دانش‌آموزانی که آزمون را گذرانده‌اند نمره‌ای برابر یا پائین‌تر از سطح آزمون گرفته‌اند. نکته مهمی که باید به‌خاطر داشت اینست که چنانچه دو دانش‌آموز در دو کلاس متفاوت اما در یک میزان همسان یک نمره واحد گرفتند، رتبه درصدی آنها ممکن است در هر یک از گروه‌های آن دو فرق کند. مثال زیر این نکته را روشن خواهد ساخت.

مثال:

علی و رضا دوقلو هستند، اما کلاسهای آن دو باهم فرق می‌کند. اخیراً هر دو در آزمون ریاضی نمره ۸۰ گرفته‌اند. نمرات سایر دانش‌آموزان در کلاسهایشان بترتیب زیر می‌باشد:

Ali's Class: ۶۴ و ۶۷ و ۷۳ و ۷۴ و ۷۷ و ۷۷ و ۷۸ و ۷۸ و ۷۹ و ۸۰ و ۸۰ و ۸۰ و ۸۲

Reza's Class: ۴۳ و ۶۵ و ۶۸ و ۷۳ و ۷۵ و ۷۶ و ۷۶ و ۷۷ و ۷۹ و ۸۰ و ۸۰ و ۸۰ و ۸۰ و ۸۵ و ۸۶

رتبه درصدی علی: $\frac{13}{17} \times 100 = 76\%$ $[F(80)] = 13$

رتبه درصدی رضا: $\frac{13}{20} \times 100 = 65\%$ $[F(80)] = 13$

مثال فوق نشان می‌دهد که رتبه درصدی نمره یک فرد اغلب مفیدتر از خود آن نمره خاص است. گرچه هم علی و هم رضا نمره ۸۰ دارند، رتبه درصدی علی بطور قابل توجهی بالاتر است. اگر ما بپذیریم که میزان رقابت در کلاسهای هر دو برابر است. این امر نشان می‌دهد که درس خواندن علی از رضا وقتی با بقیه دانش‌آموزان کلاسهای هر دو مقایسه می‌شود بهتر است. در پایان اطلاعات مربوط به تنظیم داده‌ها در جدول ۴-۵ بمنظور رجوع سریع بطور اختصار آمده است.

جدول ۴-۵ تنظیم نمودار داده‌ها

نمره	فراوانی (F)	فراوانی نسبی	درصد	فراوانی تراکمی (F)	درصد
۸۰	۵	۰/۰۸	۸	۶۰	۱۰۰
۷۰	۱۵	۰/۲۵	۲۵	۵۵	۹۲
۶۰	۲۰	۰/۳۴	۳۴	۴۰	۶۷
۵۰	۱۵	۰/۲۵	۲۵	۲۰	۳۳
۴۰	۵	۰/۰۸	۸	۵	۸

۴-۴ تنظیم گرافیکی داده‌ها

بدون تردید شیوه گرافیکی، مکمل بالارزشی برای تنظیم داده‌ها باختصار و تحلیل آماری است. یک نمودار گرافیکی یا چارت توجه خواننده را هم بخود جلب میکند. یک چنین خواننده‌ای که تقریباً بطور گذری مطلبی را میخواند، حتماً توجه کمی به مطالب چاپی معمولی در یک گزارش تحقیقی نشان میدهد و طبیعی است که جداول با انبوهی اطلاعات تلنبار شده چندان رغبتی برنیاگزیزد. با این همه، تصاویر و عکسهای ضمیمه مطمئناً نگاههای وی را جذب میکند. و همین امر امکان میدهد تا تمامی مبحث را مرور کند.

یک نمودار همیشه شیوه‌ای مؤثر و مفید در روشن ساختن یک مطلب است. اغلب نموداری کوچک خیلی واضحت‌تر از چندین پاراگراف پیرامون یک موضوع سخن میگوید. همیشه صحبت از این است که عکس‌ها زبان حال خودشان هستند. آمارها در واقع بیشتر سکوت کرده‌اند و ساکتند، جداول زبانشان بسته است. در این میان تنها نمودارها هستند که پیامشان را بصدای بلند فریاد میکنند. داده‌های معمولی بصورت ارقام کاملاً انتزاعی؛ اینها بگونه‌ای مبهم رساننده مفاهیمند و بزور به مغز خیلی‌ها فرومیروند. تصویر یا نمودار جلوه تماشائی‌تری از داده‌ها را بنمایش میگذارد.

توضیح اسپسر^۱ درباره عملکرد نمودار پایان درخوری برای این درآمد می‌باشد:

در روزهای اخیر، وقتی آموزش بصری در تمامی سطوح نه فقط یک کمک، بلکه نیز شالوده یادگیری قرار گرفته است. توجه ما بیش از پیش به امکانات تقریباً نامحدود در این زمینه معطوف گردیده است. چشمان آمار نوشتاری را جذب میکنند، اما مغز تنها باهستگی پیام پنهان در پشت کلمات و اعداد را درمیابد. با این تفصیل نمودار دقیق آن پیام را بسادگی و باختصار بیان میکند. مضمون بدانگونه که از زیر آمده از متن آن بروشنی برمیآید. متن تنها.

^۱ - Spear

۱- امکان درک بهتر داده‌ها به نسبت مطالب مکتوب

۲- تحلیل و بررسی موضوع مؤثرتر از متون مکتوب

۳- مطابقت دادن

این سه مضمون گرافیکی ضمن طراحی دقیق و آشنائی فرد با کارکردهای انواع نمودار و مدیوم‌های تصویری می‌تواند محقق گردد. شش نکته زیر شالوده طرح و تحول در ارائه نمودار گرافیکی است که داده‌های آماری را با وضوح بیشتری توضیح خواهد داد.

۱- پیام مهم داده‌ها را برساند

۲- با انواع چارت آشنا و انتخاب دقیق و درست داشته باشد.

۳- به سطح سواد مخاطبین خود اهمیت دهد. از وسایل کمک آموزشی بصری مناسب استفاده کند.

۴- آموزه‌های منطقی و مستدلی در خانه‌های جدول ارائه دهد.

۵- حد و حدود و ظرفیت خانه‌ها را تشخیص دهد.

۶- نتایج مؤثر و مفید را تشخیص دهد.

۱-۴-۴- ترسیم توزیع فراوانی

ارائه تصویر خیلی روشنی از داده‌ها از طریق روش معمولی توزیع فراوانی ممکن نیست. اما معمولاً سه شیوه ارائه توزیع نمرات بگونه‌ای گرافیکی ممکن است:

نمودار ستونی، چهار وجهی و فراوانی چند وجهی

۱-۴-۱-۴- نمودار ستونی

نمودار ستونی عموماً بمنظور توضیح گرافیکی داده‌ها بکار میرود. در این نمودار، از ستون‌های عمودی استفاده میشود. ارتفاع هر ستون نشاندهنده شمار نفرات می‌باشد؛ یعنی فراوانی آن کلاس، برای تنظیم یک چنین نمودارهایی رعایت نکات زیر توصیه می‌شود.

۱- نخست دو خط عمود برهم (یکی افقی و دیگری عمودی) رسم کنید.

۲- نمرات دانش‌آموزان را روی خط افقی و فراوانی هر نمره‌ای را روی خط عمودی بنویسید.

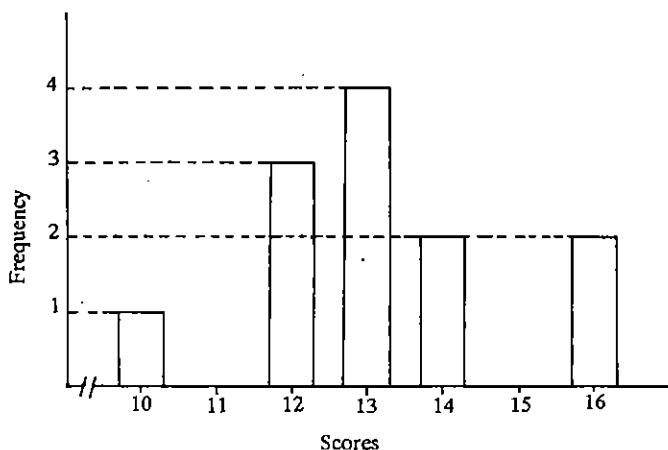
۳- ستون‌ها را در مقابل این نقاط فراوانی رسم کنید.

مثال

داده‌های زیر نتایج آزمون جایگذاری یک گروه از دانش‌آموزان می‌باشد.

۱۰ ۱۲ ۱۲ ۱۲ ۱۳ ۱۳ ۱۳ ۱۳ ۱۴ ۱۴ ۱۶ ۱۶

نمرات (x)	F
۱۳	۲
۱۴	۲
۱۳	۴
۱۲	۳
۱۰	۱



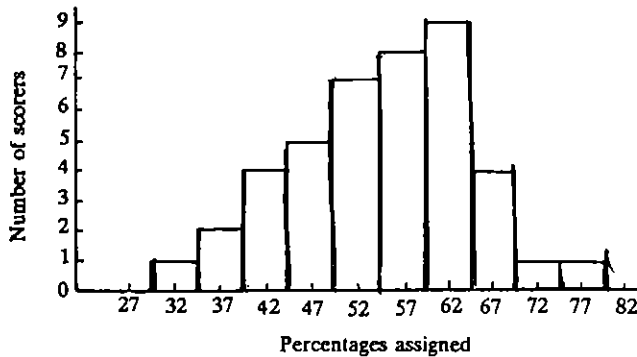
شکل ۱-۴- نمودار میله‌ای نشان‌دهنده فراوانی نمرات

۲-۱-۴- توزیع بطریق چهار وجهی

نمودار چهار وجهی عبارت است از یک تعدادی ستون که قاعده ستون‌ها نشانه نمرات کلاسها و ارتفاع ستون‌ها نشانه تعداد موارد می‌باشد. شکل ۲-۴ نمودار چهار وجهی نمرات آزمون واژگان انگلیسی ۴۲ دانش‌آموز را به‌نمایش می‌گذارد. بترتیب زیر:

نمرات کلاس	فراوانی
۲۷-۳۳	۱
۳۳-۳۷	۲

از آنجا که بزرگ‌ترین فراوانی عدد ۹ در $۶۴/۵ - ۵۹/۵$ کلاس می‌باشد، ضرورتی نداشت که خط مدرج فراوانی عمودی را بالاتر از ۹ رسم کنیم. و از آنجا که نمرات از $۵۴/۵ - ۲۹/۵$ کلاس تا $۷۹/۵ - ۷۴/۵$ کلاس متغیرند، خط افقی تا همین فاصله بیشتر لازم نبود رسم شود.



شکل ۲-۴ نمودار چهار ضلعی به نشانه نمرات آزمون واژگان انگلیسی

با این وجود برای شفافیت بیشتر، ادامه خط مندرج تا یکدرجه بالاتر و پائین‌تر رسم معمول است. برای اجتناب از ارقام خیلی هموار یا خیلی صعودی یا نزولی توصیه می‌شود تا عرض چهار ضلعی‌ها یک یا دوسوم بلندی آنها باشد - بدین معنی که نسبت بلندی به عرض حدود ۳:۵ باشد.

۳-۱-۴-۴ چند وجهی‌های فراوانی

فرایند تنظیم فراوانی‌های چند وجهی شباهت خیلی زیادی به چهار وجهی‌ها دارد. در چهار وجهی‌ها اوج هر ستونی توسط یک خط افقی مشخص می‌شود، طول هر ستون نمره به‌اندازه شماره فراوانی آن کلاس در روی خط عمودی است. اما در نمودار چند وجهی نقطه وسط هر نمره ورودی و در بالای آن درست برابر با نقطه فراوانی قرار داده می‌شود، این نقاط آنگاه با خط مستقیم بهم وصل می‌شوند. نظر به اینکه نقطه صفر نقطه آغاز و فرود در روی خط عمودی فراوانی است با وصل نقاط و کامل شدن چند وجهی بالاترین نقطه فراز و پائین‌ترین فرود چند وجهی فراوانی مشخص می‌شود. شکل ۳-۴ یک نمودار چند وجهی داده‌های مشابه نمودار چهارضلعی در شکل ۳-۴ را نشان می‌دهد.

۵-۴ آمارهای توضیحی

در فصل قبل به چگونگی ترسیم مضمون جدولی یا گرافیکی مجموعه‌ای از نمرات پرداختیم غالباً یک نمودار یا جدول آماری داده‌ها چیزی بیشتر از آنچه که ما می‌خواهیم بدانیم یا مورد احتیاجمان است دربردارد، اما رساندن پیامی که دربردارد وقت‌گیر است. بطور معمول، دو تا سه مضمون یک سری نمرات در توضیح آماری بر ما معلوم می‌گردد. این مضامین (کمیت نموداری نمرات و فراوانی آنها) ممکن است براساس شاخص‌هایی نظیر آمار اختصاری بطور دلخواه قابل توضیح باشند. در دو فصلی که در پی می‌آید آماری به اختصار که به بیان کمیت نموداری یک تعداد نمرات و فراوانی آنها می‌پردازد.

۱-۵-۴ شاخص‌های گرایش محوری

بعنوان دانش‌آموز، تنها حتماً دوست دارید بدانید که نه فقط خودتان در امتحان چه کرده‌اید، بلکه بطور کلی سایر دانش‌آموزان نیز چه کرده‌اند وقتی ما درس می‌دهیم البته که دانش‌آموزان ما هم یک چنین اطلاعاتی را خواستارند. شاخص‌هایی نظیر گرایش محوری، وضعیت نمرات، عدد واسط و میانگین بتمامی در خدمت این‌گونه اطلاعاتند.

۱-۵-۴-۱ مد^۱ نمرات

نخستین شاخص که در گرایش اصلی بسادگی قابل دست‌یابی است مد نمرات می‌باشد این مد به همراهی که بیشترین فراوانی را در ردیف نمرات حائز است مربوط میشود. حالا نه اینکه با این تعریف فکر کنیم که هر مجموعه از ارقام دقیقاً صاحب یک مد هستند.

توضیحات

۱- وقتی که تمامی اعداد در یک گروه با یک فراوانی یکسان تکرار میشوند، معمول است که بگوئیم که این گروه از اعداد دارای «مد» نیستند. بدین‌ترتیب در گروه اعداد (۱/۱/۴/۴/۷/۷) مدی وجود ندارد.
۲- زمانی که دو عدد هم‌جوار دارای فراوانی یکسانی هستند و این فراوانی مشترک بزرگتر از فراوانی سایر اعداد است، بدین‌ترتیب مد گروه اعداد مثال (۰/۱/۱/۲/۲/۲/۳/۳/۳/۴) ۲/۵ می‌باشد.
عدد واسط (۲ د) رقم ۵۰ درصد در یک گروه از اعداد می‌باشد ۴/۵/۱/۲ این همان عددی است که شماره‌های ردیف را به دو نیمه تقسیم میکند، دو نیمه‌ای که در یک سمت عدد واسط کوچکتر و در سمت دیگر بزرگتر از آن است.

توضیحات

۱- در صورتی که داده‌ها مجموعه‌ای از شماره‌های ناجور باشند، ۱۱/۱۳/۱۸/۱۹/۲۰ عدد واسط همان شماره وسطی ردیف اعداد می‌باشد. $Md = 18$
۲- چنانچه داده یک چنین ارقامی باشد، مثل گروه ۴/۹/۱۳/۱۴ عدد واسط مجموع دو عدد وسط گروه اعداد تقسیم بر دو می‌باشد. مثلاً

$$Md = (9+13) / 2 = 11$$

۲-۵-۴ میانگین

میانگین توزیع عموماً بعنوان میانگین جبری دانسته میشود. عبارت میانگین نمرات نزد دانش‌آموزان عبارت آشنائی است و بدین‌ترتیب محاسبه میشود. حاصل کل نمرات تقسیم بر تعداد نمرات. فرمول آن چنین است:

$$\text{میانگین } X = \frac{\sum x}{N} \quad \text{جمع نمرات / تعداد نمرات}$$

در حالیکه

میانگین $X =$ مقدار $\Sigma =$ نمرات افراد در توزیع $X =$ تعداد نمرات $N =$

مثال

$$\begin{array}{r} X \\ X_1 = 6 \\ X_2 = 5 \\ X_3 = 4 \\ X_4 = 3 \\ X_5 = 2 \\ X_6 = 1 \\ \hline \Sigma x = 21 \\ N = 6 \\ \hline \bar{x} = \frac{21}{6} = 3.5 \end{array}$$

میانگین مضمون جالبی دارد. چنانچه ما عدد میانگین را از هر یک از نمرات تفریق کنیم حاصل تفریق عددی است (D) که میتواند منفی یا مثبت باشد. اگر ما قرار بود که نمره D برای هر یک از نمرات N در مجموع نمرات بیابیم، مقدار نمره D برای مجموعه اعداد N صفر خواهد بود.

جدول ۴-۶

 $x = 2$ و $N = 5$ نمرات (۰/۱/۱/۳/۵)

نمره	میانگین	نمره انحراف
0	2	-2
1	2	1
1	2	-1
3	2	1
5	2	3
$\Sigma D = 0$		

یکی دیگر از خصوصیات میانگین مربوط است به نمرات D، مقدار مربع D نمرات حاصل از میانگین جبری آنها کمتر از مقدار مربع D کمتر از هر عدد دیگری است. میانگین X است مثلاً مقدار مربع D اعداد ۰/۱/۱/۳/۵ مربوط به میانگین این است:

$$(0-2)^2 = 4$$

$$(1-2)^2 = 1$$

$$(1-2)^2 = 1$$

$$(3-2)^2 = 1$$

$$(5-2)^2 = 9$$

$$\Sigma D^2 = 16$$

مثلاً مقدار مربع D اعداد ۰/۱/۱/۳/۵ مربوط به عدد ۱ بزرگتر از آنچه که مربوط به عدد ۲ می‌باشد است که در این مورد عدد ۲۱ است.

$$(0-1)^2 = 1$$

$$(1-1)^2 = 0$$

$$(1-1)^2 = 0$$

$$(3-1)^2 = 4$$

$$(5-1)^2 = 16$$

$$\Sigma D^2 = 21$$

۲-۵-۴ شاخص‌های مربوط به متغیرها (تغییرات)

گرچه مد، واسط و میانگین اعداد، در کار تحلیل مجموعه‌ای از داده‌ها خیلی مفیدند اما استفاده از آنها به تنهایی هم معایبی دربردارد. این شاخص‌ها تنها متوسط توزیع را تعیین میکنند. در پاره‌ای شرایط تعیین وسط برای ارائه تصویری منطقی از داده‌ها کافی نمی‌نماید. ما نیازمند روش‌های دیگری برای بررسی تغییرات هستیم. عبارتی، تحلیل و بررسی تفاوت میان اصطلاحات مقوله توزیع. در این فصل ما به بحث پیرامون پاره‌های روش‌های متداول مورد استفاده در امر محاسبه تغییرات خواهیم پرداخت. این روش‌ها عبارتند از دامنه تغییرات، واریانس و SD یا (Standard Deviation)

۱-۲-۵-۴ دامنه

دامنه تغییرات یک مجموعه اعداد عبارت است از تفاوت بین بزرگترین عدد و کوچکترین در توزیع. در اینجا بگذارید حسین را در نظر بگیریم که علاقمند به تبیین بهترین زمان برای آزمون واژگان است. ضمن یک ماه او واژگان را خواند و صبحها امتحان داد، و طی هفته دوم به همین ترتیب. اما امتحان در شب بود: تعداد کلماتی که او هر روز می‌آموخت براساس آزمون بدین قرار می‌باشد.

	تعداد کلمات آموخته شده .
آزمون در صبح	۱۵/۲۶/۳۰/۳۹/۴۵
آزمون در شب	۲۹/۳۰/۳۱/۳۲/۳۳

در هر یک مورد، میانگین کلماتی که یاد گرفت ۳۱ کلمه بود، اما پس کی بهتر یاد میگرفت؟ زمانی که صبحها مطالعه میکرد، تعداد کلماتی که فرامیگرفت از ۱۵ تا ۴۵ متغیر بود. بنابراین دامنه تغییرات در صبح $30 = 45 - 15$ می‌باشد. وقتی که بهنگام شب درس میخواند، تعداد واژگان یاد گرفته بین ۲۹ و ۳۳ بود. پس حدود تغییر در شب $4 = 33 - 29$ بود. از این مقایسه میتوانیم بفهمیم که آموختن کلمات توسط وی در شب به نسبت صبح از آهنگ متعادل تری برخوردار بوده است.

دامنه تغییرات تا اینجا ساده‌ترین معیار برای ارزیابی تغییرات بوده است. زیرا که تنها به دو مقدار عددی نیاز هست. با این همه استفاده از آن چندان متداول نیست زیرا که درباره چگونگی تغییرات در سایر موارد حرفی به میان نمی‌آید. بعلاوه اگر یک نمره فوق‌العاده در یک توزیع بیاید، مسئله پراکنش یا حدود تغییرات خیلی بزرگ است. اگر ما این نمره فوق‌العاده را حذف کنیم، حدود پراکنش ممکن است کاملاً کوچک گردد. بدلیل همین نقص‌ها معیارهای دیگری برای اندازه‌گیری تغییرات نظیر واریانس و معیار انحراف (کیبش) معمولاً بکار می‌آید.

۲-۲-۵-۴ واریانس و SD^۱

برای اندازه‌گیری واریانس (تغییرات) یک سری اعداد، ما نخست میانگین نمرات را بدست می‌آوریم. سپس آنرا از هر یک از اعداد تفریق و حاصل را در دو ضرب و سرانجام میانگین همه این مربعها را حساب میکنیم. حاصل واریانس نمرات خوانده میشود. حالا اگر ریشه مربع این تغییرات را بدست بیاوریم عدد Deviation اعداد را یافته‌ایم.

حدود تغییرات یک سری اعداد آیت N میانگین مربعات تفاوت‌های اعداد نسبت به میانگین می‌باشد. اگر $\bar{x}$ میانگین باشد. پس:

$$Variance = \frac{\sum (x - \bar{x})^2}{N - 1}$$

عدد Deviation یک سری اعداد ریشه مربع مثبت واریانس می‌باشد. بدین ترتیب

$$\text{Standard deviation} = \sqrt{\text{Variance}}$$

$$S = \frac{\sqrt{\sum (x - \bar{x})^2}}{N - 1}$$

بگذارید مباحث قبلی را با محاسبه Variance و Deviation معمولی یک مجموعه نمره در اینجا تصویر کنیم. حسین آزمون واژگان را گذرانده است. از آنجا که میانگین X ۳۱ است، بنابراین میتوان این محاسبه را طبق جدول ۷-۴ تنظیم کرد.

¹ - standard deviation

جدول ۷-۴ محاسبه واریانس نمرات اکتسابی بهنگام صبح

(x)	تفاوت از میانگین (x - $\bar{x}$)	مربع تفاوت (x - $\bar{x}$) ^۲
۱۵	۱۵ - ۳۱ = ۱۶	(-۱۶) ^۲ = ۲۵۶
۲۶	۲۶ - ۳۱ = ۵	(-۵) ^۲ = ۲۵
۳۰	۳۰ - ۳۱ = -۱	(-۱) ^۲ = ۱
۳۹	۳۹ - ۳۱ = ۸	(۸) ^۲ = ۶۴
۴۵	۴۵ - ۳۱ = ۱۴	(۱۴) ^۲ = ۱۹۶

از این رو اگر حسین بهنگام صبح درس خوانده باشد، واریانس نمرات وی $542/4$ یا $135/5$ و نمره Deviation وی $\sqrt{135/5}$ یا $11/64$ می باشد.

توجه کنید که در محاسبه اندازه های این واریاسیون، ما از علائمی استفاده کرده ایم. بدین ترتیب $\bar{x}$ نشانه میانگین، $x - \bar{x}$ نشانه تفاوت هر عدد نسبت به میانگین و $(x - \bar{x})^2$ نشانه مربع تفاوت است. حالا یک مثال دیگر، محاسبه واریانس و Deviation معمول نمراتی که حسین بهنگام مطالعه در شب گرفته است روش محاسبه در جدول ۸-۴ آمده است.

جدول ۸-۴ محاسبه واریانس برای نمرات بدست آمده در شب

(x)	تفاوت از میانگین (x - $\bar{x}$)	مربع تفاوت (x - $\bar{x}$) ^۲
۲۹	۲۹ - ۳۱ = -۲	(۲) ^۲ = ۴
۳۰	۳۰ - ۳۱ = -۱	(-۱) ^۲ = ۱
۳۱	۳۱ - ۳۱ = ۰	(۰) ^۲ = ۰
۳۲	۳۲ - ۳۱ = ۱	(۱) ^۲ = ۱
۳۳	۳۳ - ۳۱ = ۲	(۲) ^۲ = ۴
	مجموع = ۰	مجموع = ۱۰

در این مورد واریانس $10/5$ یا 2 می باشد و Deviation معمول $\sqrt{2}$ یا $1/41$ است. برای همین مباحث مثال دیگری بدست می دهیم.

مثال

تعداد ساعاتی که حمید صرف درس انگلیسی خود در هفته گذشته طی روز کرد $5/3/2/1/6/4/2/1/6$ می باشد، واریانس آن را به همراه Deviation معمول بیابید.

داده‌ها را بترتیبی که در جدول ۹-۴ آمده تنظیم و محاسبه را انجام می‌دهیم.

تعداد ساعات (x)	تفاوت از میانگین (x - $\bar{x}$)	مربع تفاوت (x - $\bar{x}$) ^۲
۲	۲ - ۴ = -۲	(-۲) ^۲ = ۴
۲	۲ - ۴ = -۲	(-۲) ^۲ = ۴
۳	۳ - ۴ = -۱	(-۱) ^۲ = ۱
۴	۴ - ۴ = ۰	(۰) ^۲ = ۰
۵	۵ - ۴ = ۱	(۱) ^۲ = ۱
۶	۶ - ۴ = ۲	(۲) ^۲ = ۴
۶	۶ - ۴ = ۲	(۲) ^۲ = ۴
	مجموع = ۰	مجموع = ۱۸

مراحلی که برای محاسبه واریانس و Deviation معمول باید طی کرد بقرار زیر می‌باشد:

۱- نمرات را جمع کنید و بر N تقسیم و نمره $\bar{x}$ یا میانگین را بدست آورید.

۲- Deviation هر یک از نمرات را نسبت به میانگین حساب کنید.

$$x = x - \bar{x}$$

سپس هر Deviation را مربع سازید.

۳- عدد مربع دیوی‌ایشن را نسبت به میانگین باهم جمع کنید.

$$\sum (x - \bar{x})^2 = \sum x^2$$

۴- مقدار مربع تفاوت‌ها را بر N-۱ تقسیم کنید.

۵- چنانچه ریشه مربع واریانس را پیدا کنید، به Deviation معمول دست می‌آید.

$$S = \sqrt{\frac{\sum x^2}{N-1}} = \sqrt{\frac{\sum (x - \bar{x})^2}{N-1}}$$

۳-۲-۵- محاسبه واریانس با استفاده از نمرات خام

گرچه فرمول مربوطه در فصل پیش می‌تواند همیشه برای محاسبه واریانس مورد استفاده قرار گیرد، اما بدین معنی هم هست که در عمل این محاسبه به امری کسالت بار بدل می‌شود. بهمین دلیل، از فرمولی بمراتب مناسب‌تر بگونه زیر استفاده می‌کنیم.

$$\begin{aligned}\text{Variance} &= \frac{\sum x^2}{N} - \frac{(\sum x)^2}{N^2} \\ &= \frac{\sum x^2 - (\sum x)^2 / N}{(N-1)}\end{aligned}$$

در این فرمول، که بنام فرمول نمرات خام شناخته شده، $\sum x^2$ با معنای آنست که هر عددی ابتدا باید مربع و سپس باهم جمع شوند، اما $(\sum x)^2$ به معنای آنست که نمرات بایست اول جمع زده شوند و بعد مربع گردند.

برای نشان دادن اینکه این فرمول نظیر فرمول پیشین همان نتایج را بدست میدهد، واریانس و deviation معمول یک سری نمره طبق هر دو فرمول محاسبه خواهند شد. مطابق با جدول زیر:

جدول ۱۰-۴ - نمرات فرضی

x	x^2	$(x - \bar{x})$	$(x - \bar{x})^2$
۱	۱	$1-3 = -2$	۴
۲	۴	$2-3 = -1$	۱
۳	۹	$3-3 = 0$	۰
۴	۱۶	$4-3 = 1$	۱
۵	۲۵	$5-3 = 2$	۴

$$\begin{aligned}\sum x &= 15 & \sum x^2 &= 55 & (x - \bar{x})^2 &= 10 \\ \bar{x} &= 15/5 = 3\end{aligned}$$

با استفاده از مقادیر فوق در داده‌های جدول محاسبات زیر انجام میشود.

$$\sum x = 1 + 2 + 3 + 4 + 5 = 15$$

$$\sum x^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2 = 55$$

$$(\sum x)^2 = (15)^2 = 225$$

اکنون ما میتوانیم از این فرمول برای محاسبه واریانس داده‌های جدول فوق استفاده کنیم.

$$\begin{aligned}V &= \frac{\sum x^2}{N} - \frac{(\sum x)^2}{N^2} &= \frac{55 - 225/5}{4} \\ &= \frac{\sum x^2 - (\sum x)^2 / N}{(N-1)} &= \frac{55 - 45}{4} \\ &= \frac{55 - (15)^2 / 5}{5-1} &= \frac{10}{4} \\ & &= 2/5\end{aligned}$$

اینک اگر با استفاده از فرمول تعریف شده واریانس را محاسبه کنید و نتایج حاصله را مقایسه نمائید. می‌بینید که پاسخ‌ها یکسانند.

$$V = \frac{\Sigma(x - \bar{x})^2}{N - 1} = \frac{10}{5 - 1} = \frac{10}{4} = 2.5$$

بهر صورت، به جواب رسیدن طبق فرمول نمره خام بطور قابل ملاحظه‌ای ساده‌تر مینماید. محاسن استفاده از فرمول نمرات خام این است که دیگر نیازی به تفریق کردن نمره میانگین از هر یک از نمرات در توزیع وجود ندارد.

ممکن است که اینطور بنظر برسد که محاسبه SD یک عدد بیهوده و پرپیچ و خم است. SD در واقع درباره اندازه پراکنش نمرات در یک فراوانی صحبت میکند. با مقایسه SD گروه‌های مختلف در خواهیم یافت که تا چه حد این گروه‌ها همسان و برابرند. تصور کنید که از شما خواسته شده تا در کلاس سوم راهنمایی (مرحله اول دبیرستان) به تدریس زبان انگلیسی بپردازید. ۶ کلاس یکسان وجود دارد و شما بایست یکی از کلاس‌ها را انتخاب کنید. معیار شما برای انتخاب چه خواهد بود؟ آیا نمرات خام انگلیسی سال گذشته بچه‌ها را انتخاب خواهید کرد؟ نمره میانگین در هر گروه چه خواهد بود؟ نه نمرات خام و نه نمره میانگین هیچ کدام اطلاعاتی کافی و قابل اعتماد برای انتخاب گروهی همسان در اختیار نخواهد گذاشت. معیار SD معتبرین شاخص آماری در این زمینه می‌باشد.

۶-۴ همبستگی

تا اینجا بسیاری از روش‌های تحلیل آماری یک متغیر واحد را مورد بحث قرار داده‌ایم. با این وجود، زمانی که با مسائل آماری مربوط به آموزش، روانشناسی، جامعه‌شناسی و از این قبیل سروکار داریم، معمولاً علاقمندیم که بدانیم آیا رابطه‌ای بین دو یا چند متغیر وجود دارد یا نه. مثلاً اگر مسئولان دانشگاه یک آزمون تشخیص استعداد حرفه‌ای برای ۱۰۰۰ نفر از داوطلبان جدید برگزار کنند، ممکن است علاقمند باشند که بدانند که آیا بلحاظ استعداد رابطه‌ای بین نمرات ریاضی و حرفه‌وفن وجود دارد یا نه. آیا دانش‌آموزانی که نمره تست ریاضی‌شان عالی است در تست حرفه‌وفن هم همینطورند؟ یا بسخن دیگر آیا دانش‌آموزی که نمره تست ریاضی‌اش ضعیف است در تست حرفه‌وفن هم ضعیف عمل میکند؟ بهمین‌سان مسئولان دانشگاه میخواهند بدانند که آیا رابطه‌ای بین معدل دوره دبیرستانی و قبولی در دانشگاه وجود دارد.

سئوالاتی از این دست وقتی پیش می‌آید که ما متغیرهای بسیاری در اختیار داریم و علاقمند به دانستن چگونگی رابطه بین این مجموعه نمرات هستیم. در این فصل یاد خواهیم گرفت که چگونه عددی را بدست بیاوریم که تعیین‌کننده رابطه بین دو دسته از نمرات است. این عدد بنام ضریب

همبستگی خواننده میشود. ضریب همبستگی ضمناً میتواند تعیین کننده اعتبار و اعتماد به یک آزمون هم تلقی شود.

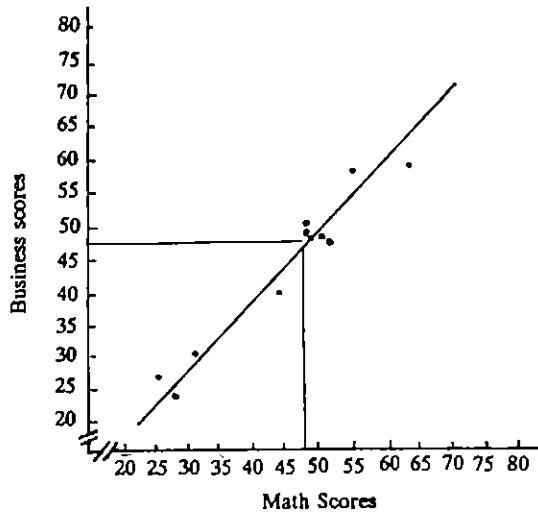
۱-۶-۴ بستگی خطی

برای درک مفهوم ضریب بستگی از موردی تجربی کمک میگیریم. بهمین منظور بگذارید مشاور راهنمایی که بتازگی نمرات تست استعداد ده دانش آموز (جدول ۱۱-۴) بدستش رسیده را در نظر آریم. مشاور هم ممکن است علاقمند به این باشد که آیا بستگی ای میان این دسته از نمرات هست. برای مثال، وی شاید بخواهد بداند که آیا دانش آموزی که نمرهش در درس ریاضی عالی است در تست مهارت های حرفه ای هم همینطور است. او به کمک یک نمودار تفصیلی این وضعیت را روشن میسازد.

برای تنظیم یک چنین نموداری، ما دو خط ترسیم میکنیم: یک خط عمودی و یکی افقی. در روی خط افقی نمرات ریاضی را نشان می دهیم. و روی خط عمودی نمرات حرفه و فن را وقتی هر دو خط مندرج گردید. از نقطه برای مشخص ساختن نمره هر دانش آموز استفاده میکنیم. نقطه ها درست بالای نمره ریاضی دانش آموز و در سمت راست نمره حرفه و فن قرار داده میشود. بدین ترتیب نقطه نمره دانش آموز A درست بالای نمره ۵۲ روی خط ریاضی و در سمت راست نمره ۴۸ خط روی حرفه و فن قرار داده میشود. بهمین سان، نقطه محل نمره دانش آموز B درست بالای نمره ۴۹ روی خط ریاضی و در سمت راست نمره ۴۹ خط حرفه و فن گذاشته میشود از همین روش برای سایر نقاط استفاده میشود.

جدول ۱۱-۴- نمرات تست استعدادهای مختلف ده دانش آموز

استعداد هنر	استعداد زبان	استعداد بازرگانی	استعداد ریاضی	دانش آموز
۲۲	۲۶	۴۸	۵۲	A
۲۳	۵۳	۴۹	۴۹	B
۵۷	۴۸	۲۷	۲۶	C
۵۴	۳۱	۲۴	۲۸	D
۱۳	۶۷	۵۹	۶۳	E
۲۰	۷۵	۴۰	۴۴	F
۹	۳۱	۷۲	۷۰	G
۵۰	۲۲	۳۱	۳۲	H

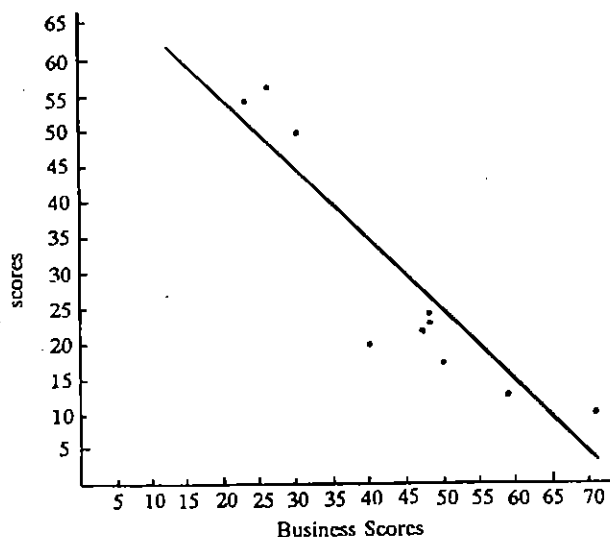


شکل ۴-۴ نمودار بار نمرات ریاضی و حرفه‌وفن

متوجه خواهید شد که این نقاط یک خط تقریباً مستقیم تشکیل می‌دهند، زمانی که این اتفاق می‌افتد می‌گوئیم که یک بستگی خطی بین دو متغیر وجود دارد.

باز هم توجه کنید که هر چه قدر نمره ریاضی بالاتر باشد، بهمان اندازه نمره حرفه‌وفن هم بالاتر است. جهت این خط از سمت چپ در پائین به سمت راست در بالا می‌باشد. وقتی این وضع پیش می‌آید، می‌گوئیم که یک بستگی مثبت بین نمرات ریاضی و حرفه‌وفن وجود دارد. این بمعنای آنست که دانش‌آموزی، با نمره ریاضی بالاتر نمره حرفه‌وفن بالاتری هم کسب می‌کند:

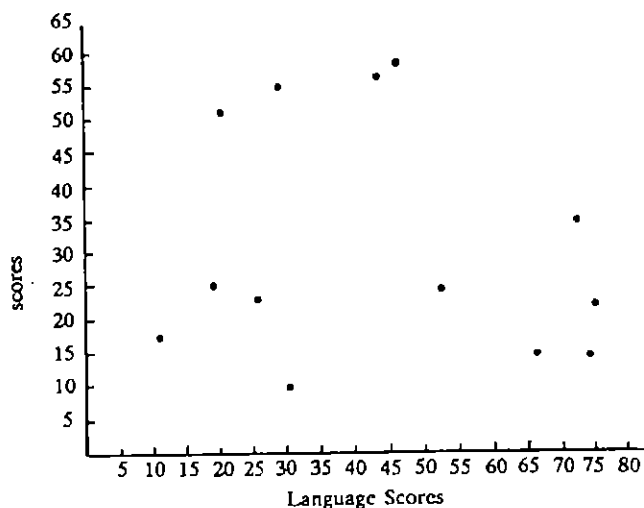
اکنون بگذارید تا نمودار بار نمرات استعداد حرفه‌وفن و هنر را چنانکه در شکل ۴-۵ ترسیم شده به تماشا بنشینیم. در این مورد شما ملاحظه خواهید کرد که هر چه نمرات حرفه‌وفن بالاتر است، نمرات هنر بهمان نسبت پائین‌تر قرار دارد. باز هم نقاط بصورت خط آرایش می‌یابند. اما این بار جهت این خط از چپ بالا به راست پائین می‌باشد. وقتی این اتفاق می‌افتد ما می‌گوئیم که یک بستگی منفی بین نمرات حرفه‌وفن و هنر وجود دارد. این امر بمعنی آنست که یک دانش‌آموز با نمره بالای حرفه‌وفن نمره هنر پائین خواهد داشت.



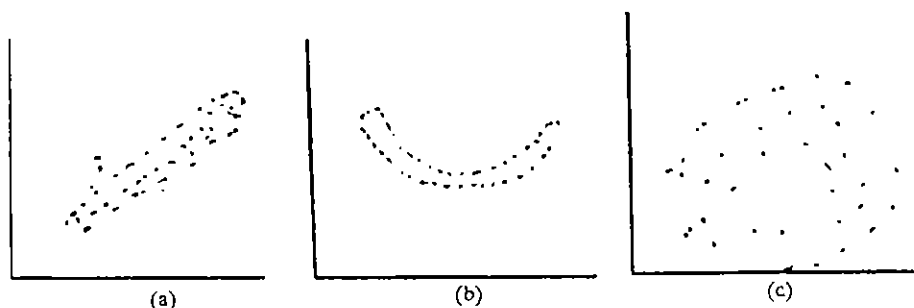
شکل ۴-۵ نمودار باز برای نمرات حرفه‌وفن و هنر

اکنون بگذارید تا نموداری برای نمرات زبان و هنر ترسیم کنیم. بدانگونه که در شکل ۴/۶ نشان داده شده، در این مورد نقاط به شکل یک خط مستقیم آرایش نمیشوند. وقتی این اتفاق میافتد باور داریم که بین نمرات زبان و هنر بستگی کمی وجود دارد یا اصلاً وجود ندارد.

گرچه ما فعلاً با نمودارهای خطی سروکار داریم، یعنی با یک خط مستقیم و یک همبستگی و با نقاطی که اشکال خمیده مختلفی را بمعرض نمایش خواهند گذاشت. این اشکال بکمک آمار بتفصیل دیرد بررسی قرار میگیرند. چندین نمونه از نمودارهای باز در شکل ۴-۷ ارائه شده است. در این فصل ما تنها به همبستگی‌های خطی و بررسی آنها خواهیم پرداخت.



شکل ۴-۶



شکل ۴-۷

۲-۶-۴ ضریب همبستگی

وقتی که معلوم کردیم که یک همبستگی خطی بین دو متغیر وجود دارد، ممکن است علاقمند باشیم که بدانیم این همبستگی تا چه حد است. کارل پیرسون^۱ پایه‌گذار ضریب همبستگی خطی است که شدت این بستگی بین دو متغیر را نشان می‌دهد. ارزش ضریب همبستگی خطی بکمک فرمول زیر قابل محاسبه می‌باشد.

فرمول ۱

ضریب همبستگی خطی طبقه فرمول زیر تعیین می‌شود.

^۱ - Karl Pearson

$$r = \frac{N(\sum xy) - (\sum x)(\sum y)}{\sqrt{[N(\sum x^2) - (\sum x)^2][N(\sum y)^2]}}$$

جائی که

نام یکی از متغیرها = x

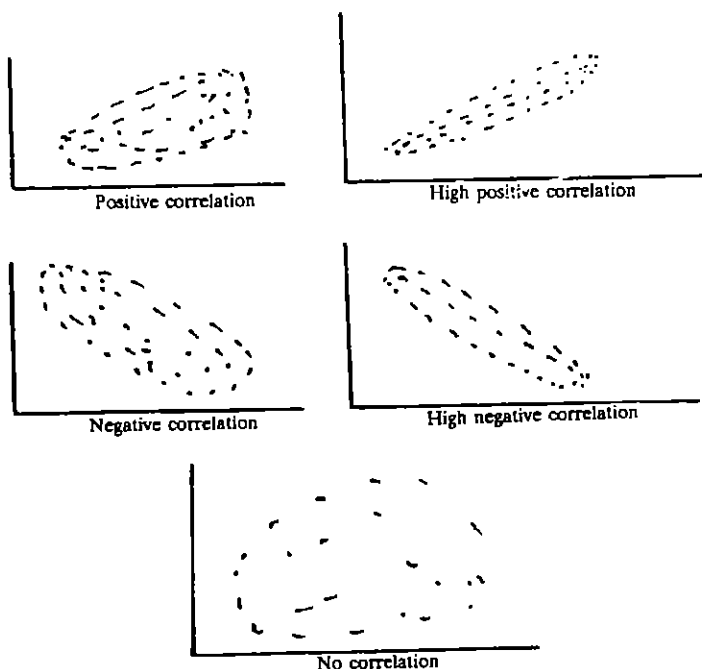
نام یکی از متغیرها = Y

N = تعداد نمرات هر دو متغیر

در این فرمول ضریب همبستگی همیشه عددی است بین -۱ و +۱ ارزش +۱ بمعنی همبستگی کامل مثبت است و منطبق با وضعیتی است که تمامی نقاط دقیقاً روی یک خط مستقیم قرار میگیرند. ارزش -۱ بمعنی همبستگی کامل منفی است یک همبستگی زمانی صعودی تلقی میشود که نزدیک به +۱ یا -۱ باشد و وقتی نزولی است که به صفر نزدیک باشد.

چنانچه ضریب همبستگی خطی صفر باشد، خواهیم گفت که همبستگی خطی وجود ندارد. این موارد در شکل ۸-۴ نشان داده شده است.

گرچه فرمول مربوطه بغرنج مینماید، استفاده از آن نسبتاً آسان است تنها علائمی که جدید الظهور است $\sum XY$ می باشد. این عدد از طریق ضرب دو عدد همسان X و Y و سپس جمع حاصل محاسبه میشود. نمونه های زیر روشهای این کار را مشخص میسازد.



شکل ۸-۴:

مثال ۱: ضریب همبستگی داده‌های زیر را بیابید.

نمره ریاضی	X	52	49	26	28	63	44	70	32	49
نمره حرفه‌وفن	Y	48	49	25	24	59	40	72	37	50

راه حل

نخست X را مصداق نمرات ریاضی و Y را مصداق نمرات حرفه‌وفن گرفتیم. بعد داده‌ها را در طرحی جدولی نظیر جدول ۴-۱۲ تنظیم می‌کنیم و بعد فرمول را بکار می‌بریم.

جدول ۴-۱۲ ضریب همبستگی جدول

X(math)	Y(business)	X ²	Y ²	XY
52	48	2704	2304	2496
49	49	2401	2401	2401
26	27	676	729	702
28	24	784	576	672
63	59	3969	3481	3717
44	40	1956	1600	1760
70	72	4900	5184	5000
32	31	1024	961	992
49	50	2401	2500	2450
51	49	2601	2401	2499
$\Sigma Y = 464$	$\Sigma Y = 449$	$\Sigma Y^2 = 23,396$	$\Sigma Y^2 = 22,137$	$\Sigma XY = 22,729$

$$\begin{aligned}
 r &= \frac{N(\Sigma X) - (\Sigma X)(\Sigma Y)}{\sqrt{[N(\Sigma X^2) - (\Sigma X)^2][N(\Sigma Y^2) - (\Sigma Y)^2]}} \\
 &= \frac{10(22,729) - (464)(449)}{\sqrt{[10(23,396) - (464)^2][10(22,137) - (449)^2]}} \\
 &= \frac{18,954}{\sqrt{(18,664)(19,769)}} \\
 &= \frac{18,954}{(136.62)(140.6)} \\
 &= \frac{18,964}{19,209.08} = 0.9867
 \end{aligned}$$

بدین ترتیب ضریب همبستگی ۰/۹۸ می‌باشد. از آنجائیکه این نمره نزدیک به ۱+ است، می‌گوئیم که یک همبستگی فوق‌العاده در کار است. نمودار باز نیز همین نتیجه را نشان میداد.

مثال ۲

ضریب همبستگی داده‌های زیر را بیابید.

نمره زبان	۲۶	۵۳	۴۸	۳۱	۶۷	۷۵	۳۱	۲۲	۱۱	۱۹
نمره هنر	۲۲	۲۳	۵۷	۵۴	۱۳	۲۰	۹	۵۹	۱۷	۲۴

راه حل

X نشانه نمرات زبان و Y نشانه نمرات هنر می‌باشند. سپس داده‌ها را در شکل جدولی، در جدول

۴-۱۳ قرار می‌دهیم.

جدول ۴-۱۳

X (Language)	Y Art	X ²	Y ²	XY
26	22	676	484	572
53	23	2809	529	1219
48	57	2304	3246	2736
31	54	961	2916	1674
67	13	4489	169	871
75	20	5625	400	1500
31	9	961	81	279
22	50	484	2500	1100
11	17	121	289	187
19	24	361	576	456
$\Sigma Y = 464$	$\Sigma Y = 449$	$\Sigma Y^2 = 23,396$	$\Sigma Y^2 = 22,137$	$\Sigma XY = 22,729$

در این آزمون از فرمول استفاده می‌کنیم. بدین ترتیب ما داریم:

$$\begin{aligned}
 r &= \frac{N(\Sigma XY) - (\Sigma X)(\Sigma Y)}{\sqrt{[N(\Sigma X^2) - (\Sigma X)^2][N(\Sigma Y^2) - (\Sigma Y)^2]}} \\
 &= \frac{10(10,594) - (383)(282)}{\sqrt{[10(18,791) - (383)^2][10(11,193) - (282)^2]}} \\
 &= \frac{18,954}{\sqrt{(41,221)(28,409)}} = \frac{-2066}{(203)(168)} \\
 &= \frac{-2066}{34104} = -0.14
 \end{aligned}$$

بدین‌سان ضریب همبستگی ۰/۴- می‌باشد، از آنجائیکه این نمره نزدیک به صفر می‌باشد، همبستگی کمی وجود دارد. شکل ۴-۶ نشانه همین حاصل است.

۳-۶-۴ بستگی به ترتیب رتبه

اگر ترتیب داده‌ها تنها براساس رتبه باشد، خیلی آسان‌تر است که از فرمول ضریب همبستگی ترتیب رتبه‌بندی $\rho(P)$ استفاده کنیم. داده‌های ورودی البته که می‌توانند به داده‌های ترتیبی بدل گردند. و این گاه گاهی انجام میشود تا امر محاسبه را آسانتر سازد. در چنین موردی ارزش عددی (P) و "I" ممکن است بشدت متفاوت باشد و تبدیل نمرات ممکن است اطلاعات بدست آمده از داده‌های اصلی را دچار افت نماید.

برای محاسبه ضریب همبستگی به ترتیب رتبه غیرپارامتریک اسپیرمن، این فرمول تقریباً بسادگی بکار می‌آید.

$$P = 1 - \frac{6\sum D^2}{N(N^2 - 1)} \quad \text{فرمول ۲}$$

در حالیکه

D = تفاوت بین دو رتبه

D^2 = مقدار مرجع تفاوت بین دو رتبه

N = مجموع تعداد رتبه‌های دو نفر

بعنوان مثال، تصور کنید که داوران جونز و اسمیت تعداد ده دانش‌آموز را در یک مسابقه حرانی رتبه‌بندی کرده باشند. ضریب همبستگی نشان داد که تا چه حدود داوری آنها بهم نزدیک است.

Pupil	Jones' Rank	Smith's rank	D	D ²
A	1	1	0	0
B	3	2	1	1
C	4	5	1	1
D	7	9	2	4
E	6	6	0	0
F	9	8	1	1
G	8	10	2	4
H	10	7	3	9
I	2	4	2	4
J	4	3	2	4
N = 10				$\sum D^2 = 28$

$$P = 1 - \frac{6(28)}{10(100 - 1)} = 1 - \frac{168}{10(99)}$$

$$P = 1 - \frac{168}{990} = 1 - 0.17$$

$$P = +0.83$$

یک ضریب همبستگی ۰/۸۳ نشان از آن دارد که داوری داوران خیلی بهم نزدیک می‌باشد.

محاسبه ضریب همبستگی بطریق اسپیرمن سریع و ساده است این روش در صورتی که داده‌ها به ترتیب رتبه‌بندی شده باشند روشی سریع و ساده است. با اینکه تعداد ارقام بین ۹ و ۳۰ باشد و ارقام بهم زیاد نباشند. معلمان این روش محاسبه را بهنگام بررسی وضعیت یک گروه دانش‌آموزی در یک کلاس واحد بعنوان موضوع تحقیق مفید می‌یابند.

۴-۶-۴ همبستگی

تصور کنید که از میان متغیرهای دوگانه متغیری (X) برای بررسی تنها نمرات ۱ یا ۰ در نظر گرفته شده باشد. و برای متغیر دیگری یعنی (Y) هر نمره‌ای که از میان نمرات ممکن باشد انتخاب شده باشد. مثلاً، اگر متغیر X شامل نمرات ۱ و ۰ باشد، متوجه می‌شویم که در مورد تمام نمرات (Y و ۰) دو نفر مفروض حاصل XY برابر با صفر میشود و دیگر اینکه در مورد تمامی نمرات (Y و ۱) حاصل XY برابر با مقادیر Y است پس: $\Sigma XY = \Sigma Y$ یا مقادیر نمرات Y, X برابر با ۱ است. آنگاه برای مقادیر دوگانه متغیر X و مقادیر متفاوت Y ما ناچاریم تا همبستگی biserial یعنی (rpb) را محاسبه کنیم. محاسبه این همبستگی در آزمون‌هایی بکار می‌آید که بخش‌های مختلف دارند، بنقل از هاچ و فرهادی (صفحه ۲۰۴ و ۱۹۸۲).

پاسخ به آیتمی واحد در یک آزمون یا درست است یا غلط. بدین ترتیب یک متغیر قراردادی دو وجهی ۱ و ۰ بوجود آمده. بنابراین نمره کل یک آزمون بعنوان یک متغیر پایدار و آن آیتم در آن آزمون بعنوان متغیر قراردادی نامیده میشود. بستگی بین هر آیتمی و بین تمامی آزمون بشیوه استفاده از همبستگی biserial قابل محاسبه می‌باشد. اگر شما بخش‌های فرعی (قرائت، واژگان، گرامر و نظایر آن در یک آزمون دارید و می‌توانید، این محاسبه را با آیتم همان درس یا هر درس دیگری انجام دهید. وقتی که این کار را انجام می‌دهید، امیدوارید ببینید که تمامی آیتمهای قرائت با نمره کل آن بستگی بیشتری داشته باشند تا با نمره کل گرامر. این نوع تحلیل و بررسی به شما خواهد گفت که آیا آیتمهای آزمونی واحد با آیتمهای دیگر درس‌ها «هم‌آهنگی دارند».

فرمولی ساده برای محاسبه rpb بقرار زیر می‌باشد.

$$rpb = \frac{x - \bar{x}}{s_x} \sqrt{\frac{p}{q}}$$

در میانگین کل آزمون دانش‌آموزانی که ۱ گرفته‌اند $\bar{x}_p$

میانگین کل آزمون دانش‌آموزانی که ۰ گرفته‌اند $\bar{x}_q$

SD تمامی نمرات توسط دانش‌آموزان s_x

نسبت افرادی که ۱ گرفته‌اند p

نسبت افرادی که ۰ گرفته‌اند $q =$

توضیحات

تست در اینجا به آیتمی اشاره دارد که پاسخ درستش نمره (۱) دارد و پاسخ غلط نمره (۰). فرمول فوق یکی از چند فرمول ابتکاری پیرسون یعنی فرمول ضریب همبستگی فوری میباشد. جدول زیر محاسبه I_{pb} براساس داده‌های گردآوری شده دایر بر رابطه جنسیت (متغیر قراردادی دوگانه) با نمرات ۱۵ دانش‌آموز، هشت پسر و هفت دختر در آزمون هشت آیتمی زبان انگلیسی آمده است. پسران بلحاظ میانگین نمره بهتری از دختران داشتند (۶۴/۲۵ در مقابل ۶۱/۱۴)، اما رابطه بین جنسیت و نمره تنها بستگی متوسطی را داشت.

تمرینات قسمت اول

۱- برای هر یک از موارد زیر روشن سازید که آیا همبستگی مثبت، منفی یا همبستگی در حد صفر را پیش‌بینی می‌کنید.

الف - عمر اتومبیل‌تان و قیمت فعلی آن

ب - طول مدت زندان و ارتکاب جنایت

س - تقلب در امتحان و استفاده از Hard drugs

د - میزان دستمزد به بانک با نرخ بیکاری

ی - تولد و مرگ افراد

ف - محدودیت‌های سرعت و حوادث رانندگی

ج - وزن و قد کودک بهنگام تولد

چ - سیگار کشیدن و تعداد مرگ و میر ناشی از آن.

آی - میزان آلودگی هوا و ابتلاء به بیماری‌های تنفسی.

ژ - افزایش وزن و بیماری حملات قلبی

۲- بمنظور تعیین رابطه بین قد و اندازه کفش تجربه‌ای انجام شده پنج نفر در اختیار گزینش شده‌اند و وزن آنها به اینچ و اندازه کفش‌ها در چارت زیر ثبت شده است.

X قد	۶۴	۶۶	۶۸	۷۱	۷۲
Y اندازه کفش	$7\frac{1}{2}$	$8\frac{1}{2}$	۸	۹	۱۱

نمودار بازی برای داده‌ها ترسیم کنید و آنگاه ضریب همبستگی را محاسبه نمایید.

۳- پروفسور X گمان می‌کند که دو تا از دانش‌آموزانش بیل و جیل باهم در امتحان دست به تقلب زده‌اند. بدین دلیل که نمرات آنها کاملاً بهم شبیه است. این دو نمرات زیر را در شش quizze اول گرفته‌اند.

Bill	X	۸۰	۶۴	۷۷	۸۵	۸۰	۹۰
Jill	Y	۸۰	۷۱	۷۵	۸۸	۸۵	۹۷

الف - نمودار بازی برای داده ترسیم و سپس ضریب همبستگی را محاسبه کنید.

ب - آیا سوءظن پروفیسور X بجاست؟

۴- یک شرکت بزرگ تولیدکننده دارو اصرار دارد تا کارگران و کارکنانش اضافه کاری بایستند. اتحادیه مدعی است که هر چه ساعات کاری بیشتر باشد بهمان اندازه ریسک کاری و حادثه ناشی از خستگی بیشتر است. اتحادیه برای تأیید این ادعا متوسط ساعات کاری یک کارگر در هفته و متوسط تعداد حوادث را براساس اطلاعات آماری گردآوری نموده است.

تعداد ساعات کار	X	۳۴	۳۶	۳۸	۴۰	۴۲	۴۴	۴۶
تعداد حوادث	Y	۲/۵	۲/۱	۲/۶	۳/۴	۶/۷	۹/۱	۵/۹

b- آیا ادعای اتحادیه کارگری درست بنظر میرسد.

قسمت B

مسائل و تمرینات

- ۱- تحولات روش‌های آماری تقریباً در سال‌های اخیر رخ داده است.
- ۲- یک سری ارزش‌ها و معیارهای معین یک مقوله واحد نام دارد.
- ۳- دو نگرش اساسی که مباحث آماری مستلزم آنند عبارتند از و
- ۴- معیار و نمره‌های اصلی داده‌های خوانده میشوند.
- ۵- جمعیت همیشه خیلی بزرگ است T F
- ۶- ما از آمار تحلیلی زمانی استفاده میکنیم که هدفمان فراتر از گروه مورد مطالعه (میرود - نمیرود)
- ۷- جهت تعمیم یک نمونه معتبر بر یک جمعیت، ضرورت دارد تا آن نمونه
- ۸- نمونه‌ای که به‌گونه‌ای گزیده شده که هر یک از اعضای دیگر جمعیت هم بتوانند مشمول آن گردند خوانده میشود؟
- ۹- چه تفاوتی بین آمار تحلیلی و آمار استنباطی وجود دارد؟
- ۱۰- نمره آیتمی که مرتب تکرار میشود نامیده میشود.
- ۱۱- نمره آیتمی که بزرگتر یا برابر نیمی دیگر از آیتمهاست و برابر با کوچکتر از نیمی دیگر است چه خوانده میشود؟
- ۱۲- نمره‌ای که مجموع نمرات آیتمهاست و بر مقدار تعداد آیتمه تقسیم میشود خوانده میشود.
- ۱۳- تا اینجا مهمترین اصطلاح از اصطلاحات سه‌گانه که فقط یک توضیح را نشان میدهد خوانده میشود.
- ۱۴- به ارقام زیر توجه کنید:

۱ ۲ ۳ ۶ ۱۲ ۱۲ ۱۷

الف) میانگین این اعداد است.

ب) مد این اعداد است.

س) واسط این اعداد است.

۱۵- اصطلاحی که اغلب نشانه حس و حال است و در اینجا بازتاب آرایش ظاهری تمامی توزیع

می‌باشد خوانده میشود (Mean / Median-mode)

۱۶- جدول را کامل کنید:

Distri bution	Mean	differences
3	_____	_____
5	_____	_____
S	_____	_____
10	_____	_____
10	_____	_____
12	_____	_____

کل تفاوت‌ها=

۱۷- زمانی که تفاوت‌ها جمع میشوند. با توجه به علامت + یا - مقدار تفاوت‌ها است.

۱۸- در هر سری نمره، مقدار از برابر است با

۱۹- در هر مجموعه نمره‌ای، چنانچه مقدار از میان پاره‌ای اعداد در توزیع با صفر برابر باشد،

آن عدد است.

۲۰- مفهوم تغییر مربوط است به رابطه میان (all,some/Nome) نمرات در فراوانی

۲۱- هرگونه تغییری در توزیع به تغییر در فراوانی منجر میگردد.

۲۲- اندازه‌گیری واریانس هر فراوانی‌ای مضامین زیر را دربردارد:

الف - برای توزیع‌هایی که در آن نمرات خیلی به هم نزدیک هستند، سنجش ارزش

(کم/زیاد) دارد. ارزش آن هنگامی که نمرات بطور وسیع پراکنده هستند ارزش آن می‌شود.

ب - ارزش سنجش ارتباط دارد با تعداد نمرات در توزیع

س - سنجش (بستگی دارد / بستگی ندارد) به میانگین

۲۳- اگر واریانس فراوانی A بزرگتر از فراوانی B باشد خواهید دانست که SD فراوانی A (larger-

smaller) از فراوانی B خواهد بود.

۲۴- در صورتی که توزیعی واریانس کوچکتری از دیگری داشته باشد، نیز کوچکتری هم

خواهد داشت.

۲۵- SD نمرات را در توزیع زیر محاسبه کنید:

۱ نمرات ۳ ۴ ۵ ۵ ۶

Standard Deviation = .

۲۶- بطور کلی برای توضیح وضعیت نمره‌ای، تمام آنچه که لازم داریم تعداد نمرات بالای آن است

۲۷- بمنظور شرح وضعیت نمره‌ای، نمرات در تمامی توزیع باید معلوم باشد.

۲۸- چنانچه، در کلاس با ۲۰ دانش‌آموز، نمره رضا پس از پنج دانش‌آموز دیگر قرار داشته باشد، ما میتوانیم بگوئیم که درصد تمامی دانش‌آموزان از رضا جلوتر هستند.

ساختار آزمون

۱-۵ مقدمه

در فصل قبل پاره‌ای اصطلاحات اساسی در آزمون زبان مورد بحث و بررسی قرار گرفت. طرز کار و هدف از یک آزمون که نقشی عمده در تهیه و تنظیم و استفاده از آن دارد به تفصیل شرح داده شد. آنگاه شکل آزمون، که همان شکل آیتمهاست مورد بحث قرار گرفت. باید به این نکته اشاره داشت که شکل و کار آزمون از هم جدا نیست. بلکه با هم پیوند درونی دارند. این بمعنای آنست که شکل خاص یک آیتم مبتنی است بر اهداف مختلف. با این همه، برحسب کار و هدف یک آزمون، پاره‌ای محدودیت‌ها بر شکل آیتم‌ها سایه می‌افکند. برای مثال، اگر قصد و هدف از یک آزمون اندازه‌گیری مهارت گفتاری است پس آزمونی نوشتاری برای این کار چندان مناسبی ندارد. بهمین خاطر است که می‌گوئیم کار هدفمند یک آزمون میتواند بر اشکال آیتمها تأثیرگذار باشد.

در ضمن باید به این نکته اشاره داشت که ترکیب آزمون یک فرایند تک بعدی نیست. بعلاوه برای تعیین شکل و کارکرد یک آزمون، موضوع و تعداد آیتمها بایست از پیش معین گردند. بویژه، برای تنظیم آن مراحل زیر باید طی شوند. نخست باید پیرامون کار و هدف آزمون تصمیم گرفت. دوم موضوع امتحان مشخص گردد. این مرحله طرح و برنامه‌ریزی است. سوم تهیه و آماده‌سازی آیتمها براساس موضوع. چهارم، آیتمها باید قبل از آزمون آزموده شوند تا معلوم گردد درصد آماری آنها به چه منوال است. و بالاخره، به آزمون باید اعتبار بخشید. بدین‌ترتیب مراحل تألیف و تنظیم یک آزمون میتواند به‌ترتیب زیر باشد.

۱- تعیین کارآمد و شکل و هدف یک آزمون

۲- طرح و برنامه‌ریزی (تعیین موضوع آزمون)

۳- تهیه و تنظیم آیتما

۴- تجدیدنظر و بررسی مجدد

۵- آزمون آیتما قبل از امتحان

۶- اعتبار بخشیدن به آن از طریق تأیید آن

هر آزمونی برای آنکه در فرایند تصمیم‌گیری بکار آید می‌باید از این مراحل گذر کند. والا نه معتبر خواهد بود و نه میشود بدان اتکاء نمود. این مراحل بلحاظ اهمیت در فصل بعد برای کمک به خواننده در جهت درک و اجرای آنها مورد بحث و بررسی بیشتر قرار خواهد گرفت.

۲-۵ چگونگی شکل و کار یک آزمون

همانطور که قبلاً اشاره شد، یک آزمون برای آنکه کار و هدفی از آن برپایند تهیه و تنظیم میشود. بهمین دلیل بخاطر تأثیری که کارکرد یک آزمون بر شکل و موضوع آن دارد، این کارکرد باید از قبل تعیین شده باشد. برای انجام این کار سه عامل لازم است که مورد توجه قرار گیرد: اختصاصات آزمون شوندگان هدف ویژه آزمون، و گستره آن.

فاکتور نخست به شناخت خصوصیات آزمون دهندگان مربوط میشود. تنظیم‌کنندگان آزمون باید مختصات وجودی یک جمعی که قرار است امتحان شوند در نظر داشته باشند. مثلاً اگر امتحان برای یک گروه نوجوان است، پس موضوع و شکل آن می‌باید با قدرت فکر و شناخت آنان متناسب باشد. بنابراین، یک آزمونی با آیتماهای تصویری خیلی مناسب‌تر از اشکال نوشتاری همین‌ها خواهد بود. یا اگر زبان آزمون شونده زبان خاص متفاوتی است، پس آزمون باید شامل آیتماهایی باشد که مبتنی است بر یافته‌های یک تحلیل افتراقی بر مبنای زبان اصلی و استراکچر آن. از طرف دیگر استفاده از یافته‌های یک تحلیل افتراقی در مورد یک گروه نامتجانس، بدلیل اینکه احتمال دارد که مفید فایده قرار نگیرد باید با احتیاط همراه باشد. علاوه بر فاکتورهای سن و زبان سیستم آموزشی‌ای که آزمون شونده بر طبق آن تحصیل کرده باید مد نظر قرار گیرد. زیرا که میزان مهارت‌های مختلف زبانی تحت‌الشعاع روش‌های آموزشی وی می‌باشد.

عامل دوم تعیین هدف آزمون می‌باشد. همانطور که قبلاً گفته شد، آزمون‌ها میتوانند در خدمت دو وظیفه عمده درآیند: امر پیش‌بینی و ارزیابی آموخته‌ها، علاوه بر تعیین کار عمده آزمون، آزمون‌سازان باید بر مبنای هدف ویژه‌ای که آزمون قرار است به‌انجام رساند به تصمیم‌گیری و سیاست‌گذاری بپردازند. چرا که کارکرد ویژه یک آزمون مثل تشخیص میزان سواد یا استعداد فرد و نظایر آن، بر شکل و موضوع آزمون اثر مستقیم دارد. مثلاً، یک آزمون تخصصی نیازمند موضوعی متفاوت با آنچه که مثلاً آزمون استعداد لازم دارد می‌باشد، با مضمون آزمون چه انجام داده‌ای با آزمون تخصصی فرق میکند، تنها به این دلیل که یکی بر مجموعه‌ای از مطالب درسی در چهارچوب یک دوره تحصیلی متکی است و یکی دیگر بی‌نیاز از این همه می‌باشد.

و سرانجام اینکه، گسترهٔ آزمون می‌باید روشن باشد. اینکه آیا آزمون مختص یک کلاس یا مدرسه است یا منطقه‌ای و فراتر از آن کشوری است، بر ترکیب آزمون اثر می‌گذارد. هر چقدر آزمون فراگیرتر باشد، میزان دقتی که باید به کار رود همراه با مقدار انرژی و زمانی که تهیه و تنظیم آزمون لازم دارد و بلحاظ آنکه هرگونه تصمیم درباره نمره دادن بر سرنوشت عده کثیری از جمع بزرگتری از دانش‌آموزان تأثیر می‌گذارد، افزایش خواهد یافت.

در کل، برای تهیه و تنظیم یک آزمون، بایست کار و شکل یک آزمون با توجه به عوامل فوق معین گردد. تهیه‌کنندگان آزمون پس از این مرحله بایستی موضوع آیت‌ها را نیز تعیین نمایند.

۳-۵ طرح و برنامه‌ریزی (تعیین موضوع آزمون)

بدانگونه که قبلاً اشاره شد، منظور از آزمون گردآوری اطلاعات کمی درباره معلومات آزمون شونده و تسلط وی در زمینه‌ای خاص می‌باشد. در همین راستا آزمون‌های زبان برای سنجش توانمندی آزمون شونده مبنی بر مهارتی خاص یا تمامی مهارت‌های زبانی طرح و برنامه‌ریزی می‌شود. بنابراین فوق‌العاده اهمیت دارد که آزمون‌کننده بر مبنای معلوماتی که می‌خواهد بسنجد تصمیم بگیرد. این بدین معنی است که موضوع آزمون باید بدقت و به تصریح مشخص گردد. گذشته از موضوع، ضمناً بر اهمیت نسبی عناصر متشکله آن صحنه گذاشته شود.

البته، تأیید بر موضوع آزمون نمیتواند مستقل از نقش و وظیفه آن باشد. و همینطور به شکل آن ربطی نداشته باشد. مثلاً مضمون آزمون جایگذاری بایستی منطبق با تست استعداد فرق کند. موضوع آزمون جایگذاری محدود به مواد درسی‌ای است که طی یک کورس معین تحصیلی آموخته می‌شود. در حالیکه مضمون یک تست استعداد هیچ ربطی به مطالب درسی مشخص ندارد. تازه شکل آیت‌های یک آزمون ممکن است بر موضوع آزمون اثر بگذارد. مثلاً یک آزمون درک مطلب در قالب آزمون چند گزینه‌ای بر آزمون‌کننده اجازه می‌دهد تا آیت‌های زیادی را تا جائیکه لازم می‌آید بگنجانند، در حالیکه در آزمون‌های توضیحی محدودیت زمان و مکان موجب می‌شود تا آزمون محدود به چند آیت جمع و جور باشد.

برای روشن ساختن مراحل مربوط به تصریح در موضوع آزمون، بد نیست تا مثالی بزنیم. تصور کنید که قرار باشد تا میزان تسلط افراد بر ساختار گرامری زبان اندازه گرفته شود. بازهم تصور کنید که همین آزمون طوری تنظیم شود که آموخته‌های دانش‌آموزان دبیرستانی را بر مبنای کتابهای درسی بخواهد بسنجد و بالاخره فکر کنید که آیت‌ها سئوالات درباره درک مطلب، چند گزینه‌ای‌ها و شکل نوشتاری باشد.

برای تعیین موضوع چنین آزمونی، نخستین گام بررسی اهداف آموزشی است. یعنی اینکه موضوع کورس طوری باید برنامه‌ریزی شود که شامل فهرستی از نکات مهم استراکچر زبان طی دوره آموزش باشد. آنگاه این موضوعات عمده به اجزای ویژه خود تقسیم شوند. حدود شروع و بسط اینها به عوامل واقعی مثل اندازه آزمون و زمان آن مربوط می‌شود. گام سوم تهیه جدول نکات گرامری است. این جدول

دارای دو سمت است در یک سمت آن موضوعات اصلی و موضوعات فرعی برحسب اهداف آموزشی فهرست میشود. در سمت دیگر آن شکل و تعداد آیتمهایی که باید درباره آن موضوعات اصلی یا موضوعات فرعی نوشته شود، بیان شده است.

منظور اصلی از جدول نکات گرامری دادن اطمینان خاطر به طراح آزمون دایر بر اینکه آزمون شامل همان نمونه‌های درسی است که ضمن کورس تدریس شده، می‌باشد. جدول نکات گرامری زیر را میتوان براساس شرح فوق تنظیم نمود.

جدول ۵-۱

اهداف آموزش محتوا	شمار آیتمها فراگیری
حروف اضافه	۳
شکل فعل	۵
شرطی‌ها	۲
نقل غیرمستقیم	۳
تفضیل	۲
کل	۱۵

این جدول، که حاکی از نکات اصلی گرامر زبان می‌باشد بدو منظور تنظیم شده، اول نشان بدهد که چه چیزهایی امتحان شوند، یعنی موضوع، دوم، تعداد سئوالات برای هر موضوع را در هر مرحله نشان دهد یعنی فراگیری همانطور که قبلاً اشاره شد، هر موضوع اصلی به موضوعات فرعی دیگری که در بطن موضوع اصلی است تقسیم میشود. جدول زیر، مطالب همه دوره را شامل نمیشود. با این همه این طبقه‌بندی اشکال فعل را تا حد ممکن نشان میدهد.

جدول ۵-۲

اهداف آموزش محتوا	شمار آیتمها فراگیری
اشکال فعل	۱
حال ساده	۱
گذشته ساده	۱
حال کامل	۱
گذشته کامل	۱
آینده	۱

البته، ضرورتی ندارد که شماره سئوالات یک شکل فعل برابر فعل دیگر باشد. یا به آیتمهایی که مثلاً فرایند روانشناختی را میسجد بهائی بیشتر از سایر فرایندها بدهیم. نکته مهم، با این همه اینست که نوع هر آیتمی باید قبل از پرداختن به تنظیم آنها روشن باشد. برای مشخص ساختن نوع آیتمهیا طراحان آزمون میباید درباره موضوع آنها تصمیم بگیرند. بکلام دیگر آنها میبایست متن زبانشناسی را که در آن عناصر معین به آزمون درآیند بروشنی مشخص سازد. این گام، که همان مرحله تهیه آیتمهیاست، بلحاظ رعایت پاره‌ای اصول و قواعد بمنظور اجتناب از تهیه آیتمهیا غلط از اهمیت بسیاری برخوردار می‌باشد. پس بهمین خاطر باقی فصل به توصیه‌هایی به مولفان آزمون می‌بر نهیه آیتمهیا دقیق و درست اختصاص خواهد یافت.

۴-۵ تهیه و تنظیم آیتمهیا

فرض این نکته که هر معلمی قادر به تألیف آیتمهیا منطقی و قابل قبول است نوعی ساده‌انگاری فوق‌العاده گمراه کننده است. البته، تجربه معلّمان یک سرمایه‌ی گرانبها در راستای کمک به کار تألیف آزمون می‌باشد. با این همه، تجربه به تنهائی کافی نیست. بدون شناخت کاملی از اصول مربوط به طراحی سئوال هر چقدر که معلّم دارای تجربه باشد باز در طرح سئوالات دچار اشکال خواهد شد بنابراین، به پاره‌ای رهنمودهای لازم در زمینه تهیه انواع سئوالات مختلف میپردازیم. لازم به یادآوری است که این رهنمودها به موضوع سئوالات درک مطلب اختصاص دارد، همچون آیتمهیا درست - غلط، چند گزینه‌ای و جایگذاری. سئوالات دیگری نظیر اشکال مختلف بیان یا درک و بیان در فصل بعد مورد بحث قرار خواهد گرفت.

۱-۴-۵ پیشنهاداتی پیرامون سئوالات درست یا غلط

آیتمهیا با اشکال درست - غلط بلحاظ فنی سئوالات دو گزینه‌ای هستند. این نوع آیتمهیا دارای عبارتی است که آزمون شونده لازم است که بخواند و سپس در مقابل علامت درست یا غلط (T/F)، (R/W) یا (C/I) (Y/N) و بالاخره (A/D) بگذارد. در هر مورد تنها دوگزینه ممکن است. بلحاظ آنکه عبارت درست - غلط از همه مصطلح‌تر است، این نوع سئوال بنام درست یا غلط معروف است. متداول‌ترین آیتمهیا درست - غلط برای سنجیدن توانائی آزمون‌دهنده دایر بر تشخیص اینکه اطلاعاتی که از طریق جمله در اختیارش گذاشته میشود آیا کافی است، می‌باشد. در نتیجه، یک چنین آیتمهیا معمولاً برای ارزیابی نتایج آموخته‌ها هستند. عبارتی، درک کردن عمده‌ترین فرایندی است که در پاسخ دادن به سئوالات درست - غلط بکار می‌آید. آنچه که در زیر می‌آید نمونه‌هایی است که موارد استفاده از این نوع سئوالات درست - غلط را در آموزش زبان نشان میدهد.

۱- درک شنیداری از طریق وسایل بصری

آزمون دهنده می‌شود

T/F a) this is a pencil

T/F b) they are books

آزمون دهنده می‌بیند

A pencil

pencils

۲- درک شنیداری با وسایل سمعی

آزمون شونده می‌شوند

(T/F) they went out for awalk. They are inside hoas?

۳- ساختار

3- Structure

the passive form of "He wrote a book." Is:

T/F A book was written by him.

۴- خواندن و درک مطلب

4- Reading cemprehen soin:

پس از خواندن متن از آزمون‌دهنده خواسته میشود تا تشخیص دهد که آیا مضمون سئوالات با توجه به اطلاعات در خود متن درست است یا نه.

سئوالات غلط یا درست، گرچه به ندرت در آزمون زبان میاید، چندان توصیه نمی‌شود. زیرا که دو دلیل دارد. اول، آنها خیلی زیاد به انتخاب شانس مربوطند، یعنی امتحان شونده پنجاه درصد شانس برای دادن پاسخ درست به کمکش میاید بدون آنکه اصلاً از نکات آزموده چیزی بداند. دوم، اینها محدود به اندازه‌گیری فعالیت‌های یادگیری فعالیت‌های یادگیری ساده در زبان هستند. فعالیت‌های پیچیده و مهم را نمیشود از طریق سئوالات درست - غلط سنجید. این دو عیب از اعتبار و اعتماد و استفاده از آیت‌های درست - غلط میکاهد.

علیرغم، معایب فوق‌الذکر، این آزمون‌های درست - غلط محاسنی هم دارند. راحت میشود تنظیم‌شان کرد و خیلی ساده بهشان نمره داد، آنها نیز میگذارند تا مؤلفان آزمون شمار زیادی از آنها را در یک تست بگنجانند. با توجه به محاسن و معایب اشاره شده آزمون‌سازان بایست در ارتباط با سئوالات درست - غلط نکات زیر را رعایت کنند:

۱- از عبارات کلی و طولانی اجتناب ورزند. این بدین دلیل است که اکثر تعمیم‌های پرطول و تفصیل همه عاقبت درست درمیایند. به نمونه‌های زیر توجه کنید:

الف - ریاست جمهور هر کشوری معمولاً انتخابی است.

ب - مردم روابط دوستانه‌ای در خیلی از جوامع دارند.

یک چنین عباراتی معمولاً درست‌ند. تا آنجا که حتی اگر آزمون شونده اطلاعی از نکات امتحانی نداشته باشد، نکته درست سؤال را درخواهد یافت.

از عباراتی که نکات بی‌ربط را اندازه میگیرند درگذرید. نگوئید حالا که موضوعات کلی توصیه نمیشود، پس بزنیم به سیم آخر و هر چه نکات بی‌فایده و بی‌اهمیت است را در قالب سئوالات درست - غلط در آزمون بگنجانیم. جمله زیر برای مثال، آیم مناسبی نیست، زیرا که اطلاعات آن آنقدر

بی‌اهمیت است که آزمون شونده را به حفظ پاره‌ای نکات جدای از درک جریان اصیل زبان در فرایند واقعی و طبیعی آن هدایت می‌سازد.

مثال The past tense of talk is talked.

از استفاده عبارات منفی بپرهیزید. کلمات منفی، مثل "NO" یا "NOT" حتماً از جانب آزمون‌دهنده، نادیده گرفته می‌شود. مخصوصاً زمانی که عباراتی با منفی، مضاعف گیج‌کننده‌اند.

مثال

The past tense of "talk" is talked.

الف) از دانش‌آموزان خواسته نشده تا درسی را حفظ کنند.

ب) به‌نظر دانش‌آموزان حفظی نبودن درس غیرقابل قبول است.

گاهی وقتها و بهر صورت استفاده از عبارات منفی غیرقابل اجتناب است. در چنین مواردی، کلمه منفی بایستی بگونه‌ای آنچنان شفاف نوشته شود که دانش‌آموزان متوجه آن شوند (با حرف بزرگ نوشتن، خط کشیدن زیر آن یا هر دو)

۴- از نوشتن جملات طولانی و پرپیچ و خم درگذرید. چنین عباراتی فراتر از نکاتی است که قرار است آزمون‌دهنده شوند. در اغلب موارد، چم و خم جمله مانع از این می‌شود که دانش‌آموزان نکات مورد آزمون را دریابند.

۵- جملات درست - غلط را تا حد ممکن بلحاظ طول، مشکل بودن و فراوانی مشابه هم تنظیم کنید. چنانچه عبارات درست بلندتر از غلط‌ها باشند، دانش‌آموزان رمز علامت زدن جمله بلندتر را بعنوان گزینه درست پیدا خواهند کرد، حتی اگر نکات مورد آزمون را اصلاً دریافته باشد.

رهنمون فوق‌الذکر بایست بمنزله پیشنهاد در نظر گرفته شود نه بعنوان قواعد و اصول. معلمان و مؤلفان آزمون بایستی از عقل و منطق در تنظیم چنین سؤالاتی سود جویند. این توصیه‌ها، بهر حال به کمک آنان خواهد آمد تا سؤالات نسبتاً منطقی‌تری را تنظیم نمایند.

۲-۴-۵ پیشنهاداتی مبنی بر تهیه آیت‌های جور کردنی

آیت‌های جور کردنی معمولاً برای سنجش نکاتی است که مبتنی بر کنار هم نشان دادن می‌باشند. در آزمون زبان، یکی از موارد استفاده از آیت‌های جور کردنی آزمودن توانایی دانش‌آموزان از بابت درک و تشخیص کلمات مترادف و متضاد می‌باشد. در مقایسه با آیت‌های درست - غلط، آیت‌های جور کردنی نیازمند فعالیت‌های پیچیده‌تر روانی است. شکل معمول آیت‌های جور کردنی تنظیم آیت‌ها در دو ستون است و آنگاه از آزمون شونده می‌خواهند تا آیت‌های مشابه از دو ستون را باهم جور سازند.

مثال

کلمات در ستون ۱ را با متضادهای آنها در ستون ۲ جور کنید. در ستون ۱ البته کلمات بیشتری قرار داده شده.

- | | |
|--------------|------------|
| ۱ | ۲ |
| 1- construct | a) accept |
| 2- refuse | b) clarify |
| 3- approach | c) destroy |
| 4- confuse | |

در تنظیم آیتمهای جورکردنی، پاره‌ای تذکرات نیز ممکن است مفید آید.

۱- در یک آیتم جورکردنی تنها از موضوعات همسان صحبت کنید. مثلاً اگر آیتمی که قرار است معلومات واژگان را اندازه بگیرد، بایست به معنی، متضاد یا مترادف محدود شود و نه آمیزه‌ای از همه این‌ها در یک سؤال باشد. علاوه بر این آیتمی که قرار باشد واژگان را اندازه بگیرد نبایستی حرفی از ساخت گرامری بزند.

۲- سعی کنید تعداد آیتمها در هر یک از ستون‌ها برابر نباشند. چنانچه شمار آیتمها در هر دو ستون برابر هم باشد، پاسخ به آخرین آنها از آیتمهای پاسخ داده شده قابل حدس است. بهمان اندازه‌ای که تعداد آیتمها در یک ستون افزایش یابد امکان حدس زدن پاسخ کاهش میابد.

۳- روش جور کردن آیتمها را در هر دو ستون بروشنی یادآور شوید. تا آزمون دهنده بداند که آیا از خط کشیدن، ضربدر زدن، شماره نوشتن یا علائم دیگر باید برای جور کردن کلمات استفاده کند یا نه.

۴- سعی کنید تعداد لغات مختصر باشد و ستون کوتاه‌تر سمت راست قرار گیرد.

فهرست بلند کلمات موجب تمرکز بیش از اندازه در یکجا خواهدشد. در عوض فهرست کوتاه از طرف دیگر هم مناسب برای امتحان‌دهنده است و هم برای امتحان گیرنده. معمولاً تعداد کلمات در یک ستون نباید بیش‌تر از هفت باشد. بهمان ترتیبی که گفته شد ستون کوتاه‌تر سمت راست باشد تا کار امتحان شونده اول در خواندن ستون بلندتر و بعد توجه به ستون کوتاه‌تر برای یافتن کلمه مشابه تسهیل گردد.

۳-۴-۵ نظراتی پیرامون تهیه آیتمهای چند گزینه‌ای

آیتمهای چند گزینه‌ای احتمالاً از پرمصرف‌ترین نوع آیتم می‌باشند. آنها قابل استفاده برای سنجش انواع مهارت‌ها هستند. به کمک سؤالات چند گزینه‌ای میتوان حاصل یادگیری را خیلی مؤثر و مفیدتر از درست - غلط‌ها یا آیتمهای جورکردنی اندازه گرفت. این است که در شرایط معمولی استفاده از آزمون، سؤالات چند گزینه‌ای فراگیرتر از هر نوع آیتم دیگری است.

بن جملات چند گزینه‌ای میتوانند دو شکل داشته باشند. ممکن است شکل سؤالی داشته باشد که یکی از گزینه‌ها جواب آن باشد. یا عبارتی باشد که یکی از گزینه‌ها آنرا کامل کند. بدلیل استفاده گسترده از آیتمهای چند گزینه‌ای در تألیف آنها بنهایت دقت را باید بعمل آورد. توجه به نظریات زیر به یقین مفید خواهدبود.

۱- بن سؤال بایست کاملاً روشن و شفاف باشد و نکته‌ای را که قرار است امتحان شود بی‌هیچ ابهامی بیان نماید. چنانچه بن اصل مسئله را توضیح ندهد، سؤال بمنزله آیت‌های درست - غلط عمل خواهد کرد نه چند گزینه‌ای.

مثال:

He is

- a) He is one of the students passing the text.
- b) One of the students failing the text.
- c) The man who left the testing session early.
- d) The man who did not participate in the test.

یک چنین آیت‌ی بدلیل آنکه اصل مسئله را توضیح نمیدهد اغفال کننده است. بنابراین هر گزینه‌ای بعنوان یک آیت مستقل از نوع درست - غلط کامل می‌کند. جهت اصلاح چنین آیت‌هایی، باید تلاش شود تا اصل اطلاعات به بن سؤال متصل گردد:

He is the man who because he got a headache un expectedly during the testing session.

- a) passed the test.
- b) Failed the test
- c) Left the testing session early.
- d) Did not participate in the testing session.

۲- بن سؤال می‌باید بخش عمده آیت‌م را به خود اختصاص دهد. هر کلمه‌ای یا عبارتی که دلالت بر تمامی گزینه‌ها کند باید در بن آن جای گیرد.

مثال:

The person is called an author.

- a) Who writes a book.
- b) Who reviews a book.
- c) Who prints a book.
- d) Who sells a book.

یک چنین آیت‌هایی با انتقال اجزای مشترک و همانند در گزینه‌ها به درون بن قابلیت اصلاح را دارند.

۳- از خیر عبارت‌های منفی باید گذشت زیرا که به یقین از سوی امتحان شونده هم نادیده گرفته میشود. همانطور که قبلاً اشاره شد، در موارد غیرقابل اجتناب، علائم منفی یا باید با حروف بزرگ، زیر خط‌دار یا هر دو نگاشته شوند.

۴- تمام گزینه‌ها بایست بلحاظ گرامری بتنهایی و هم بلحاظ ارتباط با بن سؤال درست باشند. انحرافی‌ها بهر ترتیب که در بن جمله قرار گیرند بایست غلط بنمایند.

مثال

Example:

Last year, incoming students on the first day of school.

- | | |
|------------------|------------------|
| a) enrolled | c) will enroll |
| b) will enrolled | d) should enroll |

در این آیت، گزینه (b) در ارتباط با خود عبارت غلط است چرا که در زبان انگلیسی یک چنین عبارتی اصلاً وجود ندارد. بدین ترتیب استفاده از گزینه‌های این چینی کاملاً بی‌فایده است و از جانب امتحان شونده خود بخود نادیده گرفته میشود. علاوه بر این، استفاده از انحرافی‌های غلط امتحان شونده را در معرض شکل غلط زبان قرار میدهد که همین امر اثرات منفی خود را بر فرایند زبان آموزی دانش‌آموز باقی میگذارد. از آنجائیکه، استفاده از عبارات انحرافی هر طور که حساب کنید هیچ فایده‌ای در بر ندارد در حالیکه معایب آن روشن است، از آنها کلاً باید اجتناب کرد. مثال دیگری بیاوریم.

He entered the room after the light.

- | | |
|----------------------|------------------|
| a) turning on | c) turn on |
| b) when he turned on | d) had turned on |

در این مثال، انحرافی (b) با سایر گزینه‌ها سازگاری ندارد، زیرا که تنها گزینه انحرافی است که ضمیر ربط دارد همینطور با بن سؤال هم سازگار نیست چرا که اصلاً در زبان چنین عبارتی وجود ندارد. بدین ترتیب امتحان شونده بسادگی بر آن چشم می‌بندد.
۵- هر آیتی بایست بروشنی دارای بهترین پاسخ باشد.

Example:

When Hossein called, I the house.

- | | |
|----------------|--------------|
| a) left | c) had left |
| b) was leaving | d) have left |

در این مثال، بیشتر از یک پاسخ ممکن وجود دارد. (b) ، (c) از چنین آیت‌مهانی باید اجتناب شود.
۶- تمامی انحرافی‌ها می‌باید عباراتی موجه باشند. یعنی اینکه انحرافی‌هایی که بلحاظ منطقی به نکته‌ای که مورد آزمون قرار میگیرد اصلاً ارتباطی ندارند از طرف آزمون شده هم به سادگی نادیده گرفته میشود.

مثال

Example:

To call on someone means

- | | |
|-----------------|---------------------|
| a) to visit | c) to talk |
| b) to telephone | d) a curious person |

در این جمله نکته تست، معنی فعل است. بهر صورت انحرافی (d) یک اسم گرامری است. که ارتباط با نکته مورد آزمون ندارد، پس لازم است که کنار گذاشته شود.

۷- تمامی انحرافی‌ها بایست یک اندازه و به یک میزان مشکل باشند. یک گزینه نسبتاً طولانی پاسخ درست مینماید. در خیلی از موارد آزمون‌سازان تازه کار قادر به تهیه گزینه‌های درست و در عین حال کوتاه نیستند، بعلاوه چون مایلند که قبول کنیم که گزینه‌ای حتماً درست است، ناچارند تا از کلمات

بیشتری برای رسیدن بدین منظور استفاده کنند. یک علت اینکه چرا بلندی پاسخ در گزینه‌ها به حساب درستی آنها گذاشته می‌شود همین است. گذشته از این، اگر یک گزینه فوق‌العاده مشکل‌تر از سایر گزینه‌ها باشد، یا مورد غفلت قرار می‌گیرد یا پاسخ درست محسوب می‌شود. در هر دو حالت، یک چنین انتخابی نه بر معلومات آزمون شونده، بلکه بر حدسی شانس‌ی اتکاء دارد.

۸- استفاده از عبارت "all of the above" یا "None of the above" بعنوان گزینه توصیه نمی‌شود. از این گزینه‌ها معمولاً زمانی استفاده می‌شود که آزمون‌سازان گزینه مناسبی پیدا نمی‌کنند. استفاده از "all of the above" قویاً باید به کناری نهاده شود.

"none of the above" در هر صورت اگر قرار باشد که استفاده شود بایست گاه به گاه باشد.

Example

Abundant means

- | | |
|--------------------|---------------------|
| a) plenty | c) infercarious |
| b) a great deal of | d) all of the above |

در این آیتم، آزمون‌دهنده تلاش می‌کند تا با فرض به اینکه تنها یک پاسخ درست برای هر آیتمی وجود دارد به سئوالات پاسخ گوید. آنگاه یک دفعه متوجه می‌شود که گزینه (a) ، (b) هر دو صحیح هستند. در این مورد، حتی اگر معنی کلمه را در گزینه (c) نداند. (در واقع نیازی به دانستن آن هم ندارد) او (d) را بعنوان پاسخ درست انتخاب می‌کند. از این دو، یک آیتم چهار جوابی همچون یک آیتم سه جوابی عمل می‌کند.

در مورد "none of the above" گرچه یک چنین ضعف‌هایی وجود ندارد، این عیب را دارد که پاسخ‌های نادرست را در اختیار آزمون شونده می‌گذارد.

۹- پاسخ‌های درست می‌باید تقریباً بطور برابر اما با اختیار میان گزینه‌ها توزیع شود. توزیع پاسخ‌های درست نمی‌بایست تابع الگوی مشخصی باشد. بعلاوه، برحسب شمار گزینه‌ها، در توزیع پاسخ‌های درست میان گزینه‌ها تقریباً تناسب متعادلی باید رعایت شود.

۱۰- بن نبایستی علائم گرامری در اختیار آزمون‌دهنده قرار دهد که بتواند بدون آنکه سؤال را درک کند بدان پاسخ دهد.

Example:

Ali picked an off the three and give it to his guest.

- | | |
|-----------|--------------|
| a) apple | c) tangerine |
| b) banana | d) peach |

در این مثال، حرف تعریف "an" آزمون شونده را به پاسخ درست رهنمون می‌سازد. حتی اگر آزمون شونده معنی بن را نداند، فقط از طریق نشانه گرامری در بن، می‌تواند پاسخ درست را حدس بزند. در چنین مواردی این نشانه بایستی حذف گردد.

۱۱- بن نبایستی با جای خالی شروع شود. این توصیه از مضمون آموزش هدف‌دار منشاء می‌گیرد. براساس نظریه آموزشی رمزشناسی، فرایندهای زبانی از اطلاعات شناخته شده شروع و به سمت اطلاعات ناشناخته حرکت می‌کند. شروع یک بن از خالی بمعنای آنست که آزمون شونده بایستی از

ناشناخته حرکت و به جانب اطلاعات شناخته شده حرکت نماید. بدین ترتیب جریان اطلاعات در جهتی است که مخالف جریان عادی اطلاعات می باشد.

همانطور که قبلاً اشاره شد. رهنمود فوق‌الذکر بایست بعنوان امری پیشنهادی و نه تجویزی تلقی گردد. هدف، تسهیل کار مؤلفان آزمون می باشد. بعلاوه این امیدواری هست که این پیشنهادات پاره‌ای از نقایص بدیهی آیتمهای آزمونی فعلی را رفع سازند. بایست توجه داشت که در نظر داشتن این پیشنهادات ضرورت دارد اما در راستای تهیه آیتمهای قابل قبول کافی نیست. البته آزمون‌سازان را در راستای ارزیابی نظری آیتمها یاری خواهد رساند. برای بررسی عینی کمیت و کیفیت آیتمها آزمون سازان بایست بقیه مراحل فرایند آزمون‌سازی را طی کنند. این مراحل که عبارتند از تجدیدنظر و تشخیص خصوصیات آماری آیتمهای آزمون بشرح زیر مورد بحث قرار میگیرد.

۵- تجدیدنظر (بازبینی)

در واقع، کار تألیف هر نوع آزمون تخصصی اغلب حاصل کار گروهی است. خیلی بندرت آزمونی توسط یک شخص به تنهایی تألیف میشود، زیرا که هر فردی یقیناً در معرض اشتباه قرار دارد. اینکه آیا یک آزمون حاصل کار فرد یا گروه باشد یا نباشد توصیه میشود که آنرا به شخص ثالثی بدهند تا مورد بازبینی قرار گیرد. طی مرحله بازبینی، اشکالاتی که از چشم آزمون ساز بدور مانده مطمئناً توسط مسئول بازبینی مورد توجه قرار خواهد گرفت، آنگاه وی تعدیل یا اصلاحات لازم برای رفع اشکال را پیشنهاد خواهد کرد. البته، همچون رهنمود برای تألیف آزمون، توصیه‌های تجدیدنظر کننده هم نظری است. بنابراین گرچه این پیشنهادات برای اصلاح و تعدیل ممکن است کیفیت آزمون را بهبود بخشد، اما کافی برای تألیف یک آزمون منطقی نیست. برای آنکه آزمونی بلحاظ علمی قابل دفاع باشد، لازم است تا بلحاظ عینی مورد تجربه قرار گیرد. یک چنین بررسی‌ای در مرحله بعدی آزمون سازی، مرحله قبل از آزمون نام دارد.

تا هم اینک آزمون تازه از مراحل طرح و برنامه‌ریزی، تهیه و تنظیم و نوشتن و بازبینی کردن گذشته است، شکل و کارکرد آن، در مرحله برنامه‌ریزی و طراحی موضوع آزمون براساس پاره‌ای نظریات درباره آزمون شوندگان از پیش معین میگردد. یعنی اینکه، جمعیتی که بخاطر وی آزمونی طراحی میشود، یک عامل مهم بشمار می‌آید، زیرا که ماهیت جمعیت به پارامترهای آزمون بی‌تأثیر نیست. مثلاً ممکن است آزمونی برای یک گروه کودک، نوجوان یا بزرگسال تنظیم شده باشد. در هر موردی موضوع و همینطور شکل آیتم بایست به خصوصیات آزمون دهندگان احتمالی توجه داشته باشد. والا، تهیه آزمون برای یک جمعیت نامشخص امری غیرمنطقی و غیرسودمند است.

فکر تعیین ماهیت یک جمعیت، «که بخاطر آن آزمون طراحی میشود» بدلیل آنکه پیش آزمون بر همین فکر متکی است، از اهمیت فوق‌العاده‌ای برخوردار می باشد.

در مرحله آیتم نویسی، آیتمهای منطقی بر مبنای پیشنهادات قبلاً ارائه شده، نوشته میشود. رعایت این قواعد اصلی ضروری است، زیرا که آیتمهایی که اشکال دارند فرایند پیش آزمون را مشکل میسازند. و سرانجام اینکه در مرحله بازبینی نظریات کارشناسانه مورد توجه قرار میگیرد و پاره‌ای اصلاحات روی آیتمها انجام میشود. بازبینی اهمیت دارد زیرا که بطور غیرمستقیم به مرحله پیش آزمون کمک می‌کند. این بدین معنی است که آیتمهای اصلاح شده پیش آزمون را مفیدتر میسازد. بدین ترتیب، آیتم نویسی، و بازبینی زمینه را برای پیش‌آزمون هموار میسازد.

پیش‌آزمون بمنزله انجام آزمون تازه تنظیم شده برای گروهی از آزمون شونده‌گان با مختصاتی نظیر همان گروه اصلی است. برای مثال، چنانچه آزمونی برای فارغ‌التحصیلان و مربیان طراحی شده باشد، بایستی در مورد یک چنین گروهی هم بطور آزمایشی تست شود. یا اگر آزمونی برای کودکان است. بنابراین باید برای همین گروه سنی هم پیش آزمون گردد. والا هدفی که مد نظر است تحقق نخواهد یافت.

هدف پیش‌آزمون هدفی دوگانه است. هدف نخست، تجربه عینی ویژگیهای هر یک از آیتمهاست. این ویژگیها شامل سهولت آیتم (IF)، تفاوت آیتمها (ID) و توزیع گزینه‌ها (CD) می‌باشد. هدف دوم پیش آزمون که اعتبار بخشیدن است، تشخیص و بررسی کارکرد آیتمها بطور کلی است. در واقع، از طریق اعتبار دادن، که آخرین مرحله در فرایند آزمون‌سازی است، خصوصیات یک آزمون بعنوان یک واحد کلی معلوم میشود، این خصوصیات قابل اعتبار بودن، قابل اتکاء بودن و قابل اجرا بودن می‌باشد. برای اجتناب از بار مشکلات فنی، به خصوصیات آیتم در این فصل خواهیم پرداخت و مختصات آزمون در فصل بعد ارائه خواهد شد.

۱-۶-۵ آسان بودن یک آیتم

یکی از مهمترین خصوصیات یک آیتم تنها آسان بودن آن است. بزبان غیرفنی، سهولت آیتم مربوط است به آسانی یک آیتم. آیتم چقدر آسان بود. جواب میتواند این باشد که، خیلی یا اصلاً آسان نبود. این اصطلاحات بهر حال نسبی و نظری هستند. چقدر که، خیلی یا اصلاً؟ برای اینکه بخواهیم ضریبی از سادگی ارائه دهیم، یک تعریف خیلی فنی‌تری لازم می‌آید. سهولت آیتم، بدین معنی است که نسبت پاسخ‌های درست برای هر آیتمی چقدر است. تناسب بدین معنی است که تمامی پاسخ‌های درست می‌بایست بر تعداد کل پاسخ‌ها تقسیم گردد. این عمل از طریق فرمول زیر قابل اندازه است.

$$IF = \frac{\sum C}{N} \quad \text{وقتی که}$$

IF = Item facility

$\sum C$ = Sum of the correct responses

N = total number of responses

برای اینکه درک مضمون خصوصیات آیتم را ساده‌تر گردانیم، داده‌ای ساده در زیر ارائه شده است. تصور کنید که یک آزمون ده آیتمی به دوازده نفر داده شود. چگونگی حاصل انجام آن در جدول ۳-۵ آمده. ستون‌ها نشانه نمراتند. و شماره هر خانه‌ای نشانه پاسخ افراد به هر سؤال است. هر پاسخ درستی نمره (۱) و هر پاسخ غلطی نمره (۰) دارد. با استفاده از داده‌های نمونه، بدیهی است که از میان ۱۲ پاسخ به آیت‌های شماره ۱ هفت تا غلط است. بنابراین، فرمول این سؤال بقرار زیر محاسبه می‌شود.

$$IF_{(1)} = \frac{7}{12} = 0.58$$

بعنوان یک مثال نمونه دیگر، سهولت شماره ۹ را ملاحظه کنید.

$$IF_{(9)} = \frac{12}{12} = 1$$

جدول ۳-۵

				سوالات							
افراد	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	کل
۱	۰	۱	۱	۰	۱	۱	۰	۰	۱	۱	۶
۲	۱	۰	۱	۰	۱	۰	۱	۰	۱	۰	۵
۳	۱	۱	۱	۰	۱	۱	۱	۰	۱	۱	۸
۴	۰	۱	۰	۱	۰	۱	۱	۱	۱	۰	۶
۵	۱	۰	۰	۱	۰	۱	۱	۰	۱	۱	۶
۶	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱۰
۷	۱	۱	۱	۱	۰	۱	۰	۰	۱	۰	۶
۸	۱	۰	۱	۰	۰	۰	۰	۰	۱	۰	۳
۹	۰	۰	۰	۱	۰	۱	۰	۰	۱	۱	۴
۱۰	۰	۰	۱	۰	۰	۰	۱	۰	۱	۰	۳
۱۱	۱	۱	۱	۰	۰	۰	۰	۰	۱	۰	۴
۱۲	۰	۰	۰	۱	۱	۱	۰	۰	۱	۰	۵
Total	۷	۶	۸	۶	۵	۸	۶	۲	۱۲	۶	

این نشان می‌دهد که تمامی آزمون شوندگان به این سؤال مشخص پاسخ درست داده‌اند. به سخن دیگر، این سؤال ساده‌ای است. بدین ترتیب می‌فهمیم که نمره ماکزیم یک سئوالی وقتی که تمامی آزمون دهندگان آیتمی را درست جواب می‌دهند برابر است با (۱) و برهمن اساس مشکلتترین آیتم آنی است که هیچکس بدان پاسخ درست نمی‌دهد. یا ضریب سهولت آن برابر است با صفر (۰) یک سؤال

فوق‌العاده ساده توصیه نمی‌شود، زیرا که اطلاعات مفیدی درباره معلومات آزمون‌دهنده در اختیار نمی‌گذارد. یا سؤال نباید فوق‌العاده مشکل باشد. این بدین معنی می‌ماند که یک نمره به هر نمره فردی اضافه یا از آن تفریق کنیم. البته، در موارد خاصی، چنین سؤالاتی بالارزشند و می‌بایست در آزمون گنجانده شوند. این موارد بعداً توضیح داده خواهد شد. فعلاً کافی است که بگوئیم سؤالاتی با ضریب سهولت ۰/۶۳، خیلی آسان، و سؤالاتی با ۰/۳۷، خیلی سخت هستند. چنین آیتم‌هایی بایست از آزمون حذف گردند. با تعیین ضریب سهولت یک آیت، آزمون سازان بسادگی می‌توانند مشکل بودن یک آیت را دریابند. این کار با استفاده از فرمول زیر امکان‌پذیر است.

در پاره‌ای کتاب‌ها از ضریب مشکل بودن به جای آسان بودن آیت استفاده می‌کنند. خواننده بهر حال باید در نظر داشته باشد که آسان بودن به نسبت پاسخ‌های درست مربوط است، در حالیکه ضریب مشکل بودن به تناسب پاسخ‌های غلط، زمانی که ضریب ساده بودن مشخص می‌شود. ویژگی دیگر آیت، یعنی فرق نمایی بایست محاسبه شود.

۲-۶-۵- خاصیت فرق نمایی آیت (افتراق سوال)

یکی از هدفهای آزمون گرفتن اندازه‌گیری معلومات دانش‌آموزان مبنی بر عناصر و مهارت‌های زبانی است. به کلام دیگر، منظور از آزمون کسب اطلاعات درباره اینکه چه کسی بهتر از دیگران است، در واقع کار یک آزمون تعیین تفاوت‌های بین یک آزمون‌دهنده پر معلومات و کم معلومات است. توان ممیزی یک آزمون بستگی به قدرت فرق نمایی آیتم‌های آن دارد. باز بزبان فنی، توان تمیزگذاری یک آیت به میزانی که یک آیت فرق یک آزمون دهنده با سواد و بی‌سواد را مینمایاند مربوط می‌شود. چنانچه، مثلاً اگر تمامی دانش‌آموزان به سؤالی پاسخ درست بدهند، نه فقط نشاندهنده سادگی سؤالات بوده بلکه تمیزگذار هم نیست. بدین ترتیب، رابطه‌ای بین سهولت سؤالات و فرق نمایی آن وجود دارد. آیتمی با ضریب ساده بودن فوق‌العاده بالا یا پائین به یقین دارای توان ممیزی مطلوبی نیست. در اصطلاح فنی، خاصیت تمیزگذاری بهر صورت به ضریب آن برمیگردد که حاصل مقایسه بین کار آزمون‌دهنده پر معلومات و کم معلومات می‌باشد.

از طریق مراحل زیر می‌شود ضریب ممیزی آیت را محاسبه کرد.

۱- نمرات کل آزمون بایست بترتیب از بالاترین نمره به پائین رتبه‌بندی شود. برای این کار، داده‌ها به شکل زیر تنظیم میگردند.

۲- آزمون‌دهندگان بایستی به دو گروه برابر تقسیم گردند. بهترین راه برای این کار این است که آزمون‌دهندگان را تا نصف بشماریم. در مورد داده‌های نمودار، نصف یا پنجاه درصد، همان عدد ۶ می‌باشد. پس از بالا که می‌شماریم بعد از شخص شماره ۶، خطی باید کشیده شود. افراد بالای این عدد «گروه قوی» یا به اختصار H و زیر آن «گروه ضعیف» یا F می‌باشند.

جدول ۴-۵

افراد	آیتم‌ها										کل
	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	
۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱۰
۲	۱	۱	۱	۰	۱	۱	۱	۰	۱	۱	۸
۳	۰	۱	۱	۰	۱	۱	۰	۰	۱	۱	۶
۴	۰	۱	۰	۱	۰	۱	۱	۱	۱	۰	۶
۵	۱	۰	۰	۱	۰	۱	۱	۰	۱	۱	۶
۶	۱	۱	۱	۱	۰	۱	۰	۰	۱	۰	۶
۷	۱	۰	۱	۰	۱	۰	۱	۰	۱	۰	۵
۸	۰	۰	۰	۱	۱	۱	۰	۰	۱	۱	۵
۹	۱	۱	۱	۰	۰	۰	۰	۰	۱	۰	۴
۱۰	۰	۰	۰	۱	۰	۱	۰	۰	۱	۱	۴
۱۱	۰	۰	۱	۰	۰	۰	۱	۰	۱	۰	۳
۱۲	۱	۰	۱	۰	۰	۰	۰	۰	۱	۰	۳
Total	۷	۶	۸	۶	۵	۸	۶	۲	۱۲	۶	

۳- برای محاسبه ضریب ممیز آیتم، از فرمول زیر بایست استفاده کرد:

$$\text{Item discrimination (ID)} = \frac{CH - CL}{1/2N}$$

برای نمونه، ضریب ممیزی عدد ۶ بدین ترتیب محاسبه خواهد شد:

طبق جدول از میان همان شش نفر، شش نفر از آزمون‌دهندگان در گروه قوی‌ها به این آیتم پاسخ درست داده‌اند. بدین ترتیب وجه تمایز بین ضعیف‌ها و قوی‌ها پاسخ به همین آیتم است.

$$ID_{(6)} = \frac{6 - 2}{(1/2)(12)} = \frac{4}{6} = 0/66$$

یک نکته‌ای که می‌باید در اینجا روشن شود این است که ترسیم خط حائل گاهی وقتها ممکن است مشکل‌ساز باشد. مثلاً، تصور کنید که این خط مجاور با نمره فردی شود که با نمره بالائی یکسان باشد. در چنین صورتی تصمیمی که باید در مورد محل خط گرفته شود بعهده آزمون ساز است. او ممکن است یک یا دو نفر از افراد را حذف نماید تا بتواند خط توزیع بین دو گروه را ترسیم کند. در واقع، در مورد داده‌هایی با شمار زیاد امکان اینکه چند نفر یک نمره یکسان بگیرند زیاد است. در هر حال یکی دو نفری در حول و حوش خط حائل چندان بر وجه تمیز آیتم اثر نخواهند گذاشت.

پس از آنکه مرحله ممیزی آیتم انجام شد، آزمون‌ساز می‌بایست برای قبول یا رد فهرست ویژه‌ی ممیزی معیاری قرار دهد. برخلاف مرحله سادگی آیتم، که ضریب ایده‌آل ۵۰ درصد است در مرحله ممیزی آیتم این ضریب صددرصد است. مثلاً عدد ممیز (۱) بدین معنی است که تمامی افراد در گروه قوی پاسخ سؤال را صحیح داده‌اند و در گروه ضعیف پاسخ‌ها غلط بوده است یعنی اینکه آیتم دارای حداکثر توان ممیز بوده است. با این همه تمامی آیتم‌ها دارای یک چنین نیروئی نیستند. بنابراین، آزمون‌ساز باید درباره میزان قابل قبول نمره ممیز آیتم تصمیم بگیرد. بطور منطق، هر چقدر این نمره به یک نزدیکتر باشد، آیتم ممیزتر خواهد بود. معیاد آیتم‌هایی که عدد ممیز را بالای عددی مثل ۰/۴۰ نشان می‌دهند قابل قبول‌ند.

۳-۶-۵ توزیع گزینه‌ها

امر ساده بودن و ممیزی آیتم‌ها دو پارامتر تعیین کننده در مقبولیت یک آیتم است. بر مبنای این ویژگی‌ها است که داوری درباره قابل قبول بودن یا نبودن یک آیتم را ممکن می‌سازد. نکته‌ای که بهر حال از طریق مراحل ID یا IF برآورده نمی‌شود و حاصل مرحله توزیع است. زیرا نمراتی که برای این خصوصیات بدست آمده متکی بر نوعی کمیت درباره شمار پاسخ‌های درست است. نه مرحله سادگی آیتم و نه ممیزی آیتم هیچکدام نمی‌توانند آزمون‌ساز را در کسب اطلاعات لازم مبنی بر مناسب بودن گزینه‌ها یاری رسانند. بنابراین، سومین خصوصیت یک آیتم مرحله توزیع گزینه‌هاست که بایست به نوعی تبیین شود که موجب بهبود وضعیت آزمون هم بلحاظ کمی و هم کیفی گردد. توزیع گزینه‌ها بمعنی فراوانی گزینه‌هایی است که توسط آزمون‌دهنده برگزیده می‌شود. عبارتی از طریق توزیع گزینه‌ها آزمون‌گر سودمندی انحرافی‌ها را می‌آزماید. گاهی وقتها آیتمی با سهولت منطقی و اعداد ممیز ممکن است از توزیع گزینه‌ای بدی برخوردار باشد. اگر چنانچه یک گزینه انحرافی مورد توجه آزمون‌دهنده قرار نگیرد، بایست یا حذف گردد یا اصلاح شود. والا فایده‌ای برای آن گزینه متصور نخواهد بود. مثلاً، تصور کنید که آیتم‌های ارائه شده در زیر در مورد یک گروه یکصد نفره به آزمون گذاشته شده و توزیع گزینه‌ها بقرار طرح زیر باشد:

جدول ۵-۵

سوال	نشانه	توزیع گزینه‌ها			
		A	B	C	D
۱	B	۱۰	۶۰	۳۰	۰
۲	A	۵۰	۴۰	۵	۵
۳	D	۴۰	۱۲	۸	۴۰

علیرغم این واقعیت که تمامی آیتم‌ها بلحاظ عدد ساده بودن قابل قبولند، الگوی توزیع گزینه‌ای بدی دارند. در آیتم شماره (۱) شصت آزمون‌دهنده گزینه B را انتخاب کرده‌اند که پاسخ درست بوده

است. گزینه انحرافی A ۱۰ نفر، C ۳۰ نفر، انحرافی D بهر صورت کسی را نداشت. این بدین معنی است که کارکردش رضایت بخش نبوده است. به کلام دیگر تصور میشود که نکته آیتم سه گزینه‌ای بوده است نه چهار گزینه‌ای. بنابراین این گزینه انحرافی بایستی اصلاح گردد.

در آیتم شماره ۲ گزینه درست A می‌باشد که ۵۰ نفر آنرا برگزیده‌اند. پنجاه درصد پاسخ درست نشانه سهولت آیتمی آینده‌ال است. در هر صورت افراد کمی گزینه C یا D را برگزیده‌اند. بدین ترتیب این انحرافی‌ها بایست بررسی و اصلاح شوند. و در آیتم شماره ۳، مشکل دیگری هست که تنها از طریق توزیع گزینه‌ها قابل مشاهده است. همانطور که از جدول برمیاید، گزینه درست D است که منتخب ۴۰ نفر است. گرچه درست بهمین تعداد هم A را انتخاب کرده‌اند. این بدین معنی است که گزینه A گزینه‌ای کاملاً اغفال کننده است؛ یا آنقدر به گزینه درست نزدیک است یا بنوعی دارای اشکالی است که به چشم نمی‌آید. بدین ترتیب بر طبق توزیع گزینه‌ها آزمون ساز میتواند معایب موجود در اصل گزینه‌ها را بیاید. این ملاحظات بطور قطع آزمون ساز را به اصلاح کیفیت گزینه‌ها رهنمون خواهدشد. این اصلاح به نوبه خود مختصات آیتما را بهبود خواهدبخشید و سرانجام اینکه تمامیت آزمون را تحت تأثیر قرار خواهدداد. خصوصیات یک آزمون بعنوان یک کلیت یکپارچه که در مرحله اعتبار یافتن صاحب آن میشود در فصول بعد مورد بحث و بررسی قرار خواهدگرفت.

۷-۵ اصول عقل و منطق را رعایت کنید

هدف از پیش آزمون پی بردن به خصوصیات ویژه آیتم است اعداد مربوط به ضریب سادگی و ممیز آیتم همراه با اطلاعات مربوط توزیع گزینه‌ها بخش مهمی از اطلاعاتی است که به آزمون ساز کمک میکند تا آیتمی را بهمان حال نگه‌دارد، اصلاح کند یا بکناری بگذارد. در هر حال، وقتی یک آیتم با وضعیتی منطقی خصوصیات آیتمیک ضعیفی از خود نشان می‌دهد، آزمون ساز نمی‌بایست اصلاً به چنین خصوصیتی اتکاء کند. او بایست از حکم عقل و منطق در انتخاب گزینه‌ها (آیتما) پیروی کند. اگر آیتمی، بدین دلیل یا آن، لازم است که در آزمونی گنجانده شود، برغم اختصاصات قابل اعتنائی که ندارد، خوب میشود آنرا نگه‌داشت. آنهم بدلیل پاره‌ای عوامل که ناشناخته مانده‌اند مثل خصوصیات آزمون شونده، برنامه‌ریزی برای امتحان و از این قبیل. بدین ترتیب، اطلاعات آماری بایست بعنوان ابزاری رهنمودی و نه تعیین کننده در گنجاندن یا حذف یک آیتم بکار آید.

نکته آخری که لازم به اشاره در اینجا است این است که توصیه‌های که میشود مبنی بر اصلاح یا حذف یک آیتم خاص بر مبنای مختصات آن می‌بایست با احتیاط عمل کرد. همانطور که قبلاً اشاره شد. آیتمی با ضریب سهولت ۱ یا صفر یک آیتم بد به حساب می‌آید تنها بدین دلیل که اطلاعات مفیدی در اختیار آزمون کننده قرار نمی‌دهد. این امر زمانی مصداق دارد که نظریه رایج درباره آزمون‌ها رعایت میشود. طبق این نظریه هدف اصلی مقایسه کار دانش‌آموزان با یکدیگر و رتبه‌بندی آنهاست. بدین ترتیب، اضافه شدن یک نمره (بدلیل آسان بودن یک سؤال) یا کم شدن یک نمره (بدلیل مشکل

بودن یک سؤال) بر رتبه آزمون شونده تأثیری نمی‌گذارد. بهمین دلیل هم هست که توصیه میشود تا سؤالات خیلی راحت و خیلی سخت از متن آزمون برداشته شوند. با این حال در آزمون‌های متکی بر معیار یک چنین آیتمهایی به هر حال اطلاعات مفیدی هستند. یعنی اینکه وقتی که تمامی دانش‌آموزان به یک آیتم خاص پاسخ درست میدهند، بمعنی آنست که تدریس کاملاً با موفقیت همراه بوده است. از طرف دیگر، اگر تمامی دانش‌آموزان نتوانستند به آیتمی پاسخ درست بدهند، این مبین آنست که تدریس با موفقیت چندانی روبرو نبوده است. بنابراین، برای اینکه درباره حذف یا جذب یک آیتمی در آزمون تصمیمی گرفته شود، فرضیات نظری پیرامون بررسی نمرات بایستی مورد توجه قرار گیرند. این نظریات و فرضیات در فصول بعد مورد بحث قرار خواهند گرفت.

مشخصات یک آزمون خوب

۱-۶ مقدمه

همانطور که در فصل پیش اشاره شد، هر یک از آیتمها سنگ بنای سازه یک آزمونند. یک آزمون ممکن است شامل تنها یک آتم یا مجموعه‌ای از آیتمها باشد. بدون توجه به شمار آیتمهای یک آزمون، هر آیتمی بایست خصوصیات ویژه‌ای را دارا باشد. یعنی اینکه هر آیتمی باید از بابت ساده بودن، شاخص بودن و توزیع گزینه‌ها مطابق با معیار باشد. بر همین اساس آیتمی که اشکال دارد باید اصلاح یا کنار گذاشته شود. با حذف چنین آیتمهایی، آزمون ساز با تعداد معینی آتم مقبول و منطقی آزمون را به سرانجام خواهد رساند.

با این همه، یک سری آتم سنجیده ضرورتاً بمعنای یک آزمون خوب داشتن نیست. برای اینکه یک آزمون بعنوان یک کل، چیزی بیشتر از جمع آیتمهاست. در واقع، آزمون‌سازان باید بخاطر داشته باشند که آتم خوب لازم است اما برای اینکه کارکرد موفقیت آمیزی هم داشته باشد کافی نیست البته آیتمهای خوب در دست آزمون‌ساز همان مصالح خام مناسب است. در هر حال زمانی که این مواد خام در این مورد آیتمها، کنار هم نهاده میشوند، بخاطر اینکه در کل آزمون خوبی داشته باشیم نکاتی چند باید مورد توجه قرار گیرد. مثالی این نکته را روشن خواهد ساخت.

تصور کنید کسی بخواهد خانه‌ای بسازد. با فرض اینکه مصالحی با کیفیتی عالی در اختیار داشته باشد، مثل آجر، تیرآهن و نظایر آن. با این همه مصالحی این چنین بهر ترتیب ضمانتی دایر بر اینکه خانه‌ای خوب هم بنا شود نیست. عوامل دیگری هم، همچون طراحی بنا و شیوه‌ای که مواد باید مورد

استفاده قرار گیرند بایستی در محاسبات بنا لحاظ شود. والا، حتی با بهترین مصالح امکان ساختن بنایی بد و نامناسب همواره وجود دارد.

این قیاس میتواند به مورد آزمون سازی هم تعمیم یابد. یعنی اینکه کنار هم نهادن تعدادی آیتم خوب و پراکنده الزاماً بمفهوم یک آزمون مطلوب نیست. در اینجا هم عوامل چندی، مثل فرایند اجرا و شیوه‌های نمره دادن مهم هستند. بنابراین علاوه بر در اختیار داشتن آیتم خوب، آزمون باید دارای مختصات مشخص هم باشد. این خصوصیات عبارتند از مطمئن بودن، معتبر بودن و مجزا بودن از آنجا که این مفاهیم در فرایند ساختاری آزمون کاملاً ضرورت دارند و نیازمند تحلیل کاملاً جامعی هستند، این فصل به روشن ساختن آنها اختصاص یافته است.

۲-۶ پایایی

تمامی دانش‌آموزان بالاخره چند باری امتحان داده‌اند. پس از هر امتحان، دارای حس و حالی هستند که انگار آطور که میخواستند امتحان نداده‌اند. که از سؤالات داده شده گوئی شوکه شده‌اند. یا شاید هم شب قبل از امتحان از دلهره خوب ن خوابیده‌اند و یا اصلاً خوب درس نخوانده‌اند. مواقعی هم هست که چند تائی از آنها حداقل جلسه امتحان را با این حس رضایت ترک میکند، گوئی که بیشتر از آنچه که انتظار میرفته خوب امتحان داده‌اند شاید بخاطر آنکه قسمت‌هائی را که باید، خوانده بودند یا اینکه شانس‌ی علامت زده بودند.

با توجه به این اظهارات گوناگون، با اطمینان میشود گفت که دانش‌آموزان درباره امتحان احساساتی فکر میکنند. البته، از اظهارات فوق شاید گریزی نباشد. در هر صورت خیلی از این اظهارات مبتنی بر تجربیات شخصی هست گرچه تجربیات شخصی معیار خوبی است. اما برای نقد در زمینه آزمون‌ها نیست. بیشتر این انتقادات از سر احساسات است تا از روی عقل و منطق. تنها در پاره‌ای موارد آزمون‌ها نگاهی دقیق انداخته‌اند. بندرت بدنبال انگیزه‌ای جدای از تجربه شخص هستند. یک بخش عمده‌ای از این انگیزه که موجب می‌شود تا آزمون شونده یا آزمون‌گیر دوباره آزمون داوری عقلانی و منطقی‌ای بعمل آورند پایا بودن آزمون است.

برای تشریح مفهوم پایا بودن بزبان فنی یا غیرفنی تا حد ممکن تکرار مجدد آنچه که در مباحث پیش بیان شد خالی از فایده نیست. این احساس که کسی آنطور که میتوانسته امتحان بدهد اما نداده، بزبانی دیگر بگونه زیر میتواند بازگو شود: اگر فرصتی دوباره می‌یافت، حتماً بهتر امتحان میداد. این البته کاملاً درست مینماید. معهذا، فرد باید قبول کند که عواملی چند قبل امر پاسخگوئی شانس‌ی به سؤالات براساس حدس و گمان سطح نمرات را از آن چیزی که واقعاً باید می‌بود بالاتر برده است اما، بندرت کسی در این باره شکوه یا از خود انتقاد میکند.

اکنون، اگر فردی بارها و بارها و به تکرار مورد آزمون قرار گیرد خواهدگفت که میانگین نمرات وی در همه این امتحانها ارزیابی قابل قبولی است از آنچه که او واقعاً میداند یا دوباره آزمون حس میکند.

برحسب یک آزمون پایا نمرات یک فرد پس از چند بار امتحان شدن چندان باهم فرقی نمیکند. یعنی اینکه نمره وی هر بار همان خواهد بود. از طرف دیگر براساس یک آزمون غیرپایا نمره فرد از امتحانی به امتحان دیگر در نوسان خواهد بود. یعنی اینکه نمره وی در هر امتحان شدن ثابت نمی‌ماند. این ثابت ماندن نمره فرد با توجه به میانگین نمره وی در هر بار امتحان گرفتن مضمون اصلی مفهوم موثق بودن است. برای بررسی مفهوم موثق بودن یا ثابت بودن یک آزمون نخست یک چهارچوب نظری بایست بوجود آید. نظریه‌های مختلفی مبنی بر توضیح مفهوم پایا در اصطلاح علمی پدید آمده است. هر نظریه‌ای برپایه یک سری مفروضات بنا می‌شود که همان قاعده کلی است و می‌باید مسلم فرض شود. از میان نظریه‌های موجود، که در این کتاب مورد حمایت قرار گرفته نظریه کلاسیک در باره آزمون است. دلیل اصلی این انتخاب شیوه ساده‌ای است که مفهوم پایایی در این نظریه تشریح شده است.

برحسب این نظریه پایا بودن یا نبودن بشرح زیر توضیح داده میشود. اگر فرد یک یا دو خصوصیت وصفی یک موصوفی را مثلاً دانش عملی یا بلندبالایی را در نظر آرد می‌بیند که این دو صفت اصلاً شباهتی بهم ندارند. از طرف دیگر، امتحان شدن مکرر موصوف شخص واحد که تقریباً هیچگاه المثنی یکدیگر نیست، «پایا نبودن» نام دارد. از سوئی امتحان شدن مکرر همان موصوف همان شخص نشان از نوعی پایداری و ثبات دارد. میل به ثبات از یک امتحان تا امتحان دیگر پایا بودن است.

بعبارت دیگر، زمانی که شخصی نمره‌ای در امتحان میگیرد، در بیشتر موارد این نمره مطابق با تصور او از توانائی خودش نیست. یعنی اینکه به این دلیل یا آن دلیل اگر دوباره مورد آزمون قرار گیرد، نمره وی تغییر خواهد کرد. همانگونه که قبلاً اشاره شده این امر دلایل مختلفی دارد. بهنگام آزمون، ممکن است سردرد داشته یا عوامل مختل کننده دیگر در کار باشد. ممکن است وقت بیشتری برای مطالعه مجدد بخاطر امتحان نداشته یا شب قبل از آن بخاطر امتحان حسابی نخوابیده، یا از طرف دیگر ممکن است با حدس و گمانهائی که درباره نمره میزده روز خوشی داشته است.

تغییرات نمرات فرد غیرقابل اجتناب است. پاره‌ای از این تغییرات نشان از افزایش مستمر نمره فرد دارد. این افزایش بطور قطع از نوعی آموزش حکایت دارد. این گونه تغییرات. که قابل پیش‌بینی است، تغییرات سیستمیک خوانده می‌شود. پاره‌ای از تغییرات، بهر حال ممکن است بدلیل عواملی مثل این که روز امتحان چه حالی داشته، یا در چه وضعی امتحان داده و نظایر آن، در اکثر موارد، تغییرات تحت چنین شرایطی که گاهی اصلاً قابل پیش‌بینی نیست تغییرات غیرسیستمیک خوانده میشود. بدین ترتیب هرگاه فرد چندین بار امتحان بشود و چندین نمره از موضوع توانمندی وی ثبت گردد، آن نمرات بطور قطع با یکدیگر متفاوت خواهند بود. تغییرات از جهتی از نوع تغییرات سیستمیک و از جهت دیگر غیرسیستمیک هستند. تغییراتی که به خطا هستند مربوط میشوند به غیرقابل اطمینان بودن یک آزمون.

برای اینکه قابل اطمینان بودن یک آزمون را محک بزنیم پس لازم است تا بررسی‌ای صورت گیرد که تغییرات سیستمیک کنترل شود و بماند. تفاوت‌های نمرات آزمون که ناشی از تغییرات

غیرسیستمیک یا خطاهای عمدی و سهوی است. این بمعنی آنست که این بررسی باید درصدد تکنیک‌هائی برآید که هم تغییرات سیستمیک و هم غیر سیستمیک را توجیه سازد. توضیح شبه فنی زیر به روشن شدن مسئله کمک خواهدکرد.

تصور کنید که فردی آزمونی بدهد. از آنجا که تمامی شیوه نمره دادن به خطا است نمره‌ای که فرد از بابت آزمون دریافت میکند نمیتواند نشانه توانمندی کامل وی در زمینه‌ای خاص باشد. به کلام دیگر، نمره توانمندی وی هم صحیح و هم خطا است. چنانچه این قسمت خطا از آن پاک شود، عدد باقی مانده نمره صحیح توانمندی وی به حساب میاید. توضیح آن این است که این نمره بدون خطا «نمره واقعی» است.

این نمره واقعی تقریباً همیشه با نمره‌ای که هر فرد میگیرد فرق دارد. که همان «نمره ظاهری» است. از آنجائیکه نمره ظاهری شامل خطای محاسبه هم هست، یعنی نمره به خطا، این نمره یا بزرگتر یا کوچکتر است و در هر حال مساوی با آنچه که باید نیست. اگر مطلقاً اشتباهی در کار نباشد نمره ظاهری با نمره واقعی یکی است. بهر صورت، وقتی خطای محاسبه وجود دارد و اغلب هم قاعده چنین است، این امر به زیاد محاسبه کردن یا کم محاسبه کردن نمره میانجامد که با نمره صحیح فرق دارد. بنابراین، اگر نمره ظاهری X و نمره واقعی T و نمره به خطا E باشد، رابطه بین نمره ظاهری و نمره واقعی بدین گونه تصویر میشود.

$$X = T \text{ یا } X > T \text{ یا } X < T$$

این رابطه، در هر صورت، وقتی که نمرات واریانس‌های خودشان را پیدا میکنند رابطه‌ای حقیقی نیست. واریانس نمرات حقیقی بلحاظ آنکه نمرات واقعی پایدارند تغییری را نشان نمی‌دهند. واریانس نمرات ظاهری معهداً بدلیل حدود خطای محاسبه در نوسان قرار دارد. زیرا واریانس نمره بخطا جزء واریانس نمره ظاهری است. واریانس نمرات ظاهری همیشه بزرگتر از واریانس نمرات حقیقی است. یعنی اینکه، بزرگنمایی واریانس نمره ظاهری برابر است با بزرگی واریانس حقیقی باضافه واریانس نمره بخطا. اگر واریانس نمرات ظاهری V_x و نمره حقیقی V_t و نمره بخطا V_e باشد فرمول شماره ۱ بدین گونه مجدداً نوشته میشود:

$$(2) V_x = V_t + V_e$$

این سه واریانس بلحاظ درک مفهوم مورد اطمینان بودن در اصطلاح آماری خیلی اهمیت دارند. از این فرمول چنین برمیاید که رابطه‌ای نزدیک بین درجه خطا در محاسبه و مقدار واقعی نمره حقیقی وجود دارد. هر چقدر خطا در محاسبه بیشتر باشد بهمان اندازه برآورد نمره واقعی کوچکتر خواهدبود. از طرف دیگر وقتی، بخش کوچکی از نمره ظاهری ناشی از خطای محاسبه باشد، یعنی اینکه برآورد نمره حقیقی به اصل آن نزدیک‌تر است.

درک مضمون واریانس حقیقی، خطا و ظاهری کمک میکند تا تعریفی از مورد وثوق بودن بلحاظ فنی بدست دهیم. پایان بودن تعریفی چنین دارد؛ نسبت واریانس نمره واقعی بر واریانس نمره ظاهری. چنانچه پایا بودن مدنظر باشد، فرمول بشکل زیر نوشته میشود.

$$(3) \quad r = \frac{V_t}{V_x}$$

البته، نمره حقیقی قابل اندازه‌گیری نیست و بدین ترتیب ارزش V_t هیچگاه شناخته نمی‌آید. بنابراین، ما میتوانیم V_t ناشناخته را از طریق محاسبه زیر بدست آوریم:

$$V_x = V_t + V_e \quad \text{یا} \quad V_t = V_x - V_e$$

حالا اگر مقدار V_t را فرمول ۳ قرار دهیم، داریم:

$$(4) \quad r = \frac{V_x - V_e}{V_x}$$

از فرمول ۴ نتایجی حاصل می‌آید. اول اگر محاسبه خطا نداشته باشد، واریانس خطا عدد صفر است. $V_e = 0$ بدین ترتیب خواهیم داشت:

$$\begin{aligned} r &= \frac{V_x - V_e}{V_x} \\ &= \frac{V_x - 0}{V_x} \\ &= \frac{V_x}{V_x} = 1 \end{aligned}$$

این بدان معناست که وقتی خطائی در محاسبه وجود ندارد ضریب اطمینان برابر با عدد (۱) می‌باشد. دوم چنانچه خطای محاسبه خیلی بزرگ باشد، نمره ظاهری خیلی با نمره اصلی فاصله دارد. نمره ظاهری همه‌اش اشتباه است. در این حالت $V_x = V_e$ و خواهیم داشت:

$$\begin{aligned} r &= V_x = V_e \\ &= \frac{V_x - V_x}{V_x} \\ &= 0 \end{aligned}$$

این بمعنی آنست، وقتی که در محاسبه اشتباه بزرگی شده، ضریب اطمینان به صفر میرسد. بدین ترتیب، بزرگی ضریب اطمینان از صفر تا یک متغیر است. ضریب صفر اطمینان که ضریب حداقل است، به این معناست که تمامی واریاسیون ظاهری ناشی از خطا است. یعنی اینکه، آزمون کاملاً غیرقابل اتکاء است. از طرف دیگر ضریب اطمینان (۱) حاکی از آن است که اشتباهی در کار نیست و آزمون کاملاً قابل اتکاء است. البته این امر در واقعیت اتفاق نمی‌افتد. تمامی آزمون‌ها درجه معینی از

عدم اطمینان را نشان می‌دهند. اما هر چقدر بزرگی ضریب اطمینان به واحد نزدیک‌تر باشد، به آزمون اطمینان بیشتری خواهد بود.

باید در اینجا اشاره کرد که میزان واریانس اشتباه از ضریب اطمینان و در نهایت از وقت تصمیم‌گیری بر مبنای نمره آزمون خواهد کاست. از این رو راهی باید تا این عدم دقت مرتفع شود. البته، رفع اشتباهات در هر شرایط ممکن نیست. حتی اگر امکان هم داشته باشد، بدلیل آنکه بخش اشتباه نمره سیستمیک نیست و در وضعیت‌های مختلف فرق میکند، چندان مفید نخواهد بود. بنابراین، ضرورت دارد تا ضریب خطا در نمره دادن را در تمامی شرایط بیابیم. این ضریب خطا بنام ضریب خطای معمول (استاندارد) و به اختصار SEM خوانده میشود. SEM راه برگشت تمامی نمرات اشتباه اکتسابی براساس معیاری معین در شرایط مختلف می‌باشد. برای محاسبه ارزش عددی SEM فرمول ضریب اطمینان از طریق روش زیر قابل محاسبه است:

$$(1) \quad r = \frac{V_t}{V_x}$$

$$V_x = V_t + V_e$$

$$V_t = V_x - V_e$$

$$2) \quad \frac{V_x - V_e}{V_x}$$

یا:

$$r = \frac{V_x}{V_x} - \frac{V_e}{V_x}$$

$$(3) \quad r = 1 - \frac{V_e}{V_x}$$

$$(4) \quad V_e = V_x(1 - r)$$

با یادآوری مبحث فصل (۴) Standard deviation ریشه مربع واریانس می‌باشد. ریشه مربع فرمول شماره ۴ چنین است.

$$\sqrt{V_e} = \sqrt{V_x(1 - r)}$$

or:

$$Se = Sx\sqrt{1 - r}$$

ارزش Se یعنی Standard deviation خطاها، همانطور که قبلاً گفته شد SEM خوانده میشود.

بدین ترتیب

$$SEM = Sx\sqrt{1 - r}$$

در این فرمول، S_x دلالت بر SD نمرات ظاهری و r ضریب اطمینان می‌باشد. همانطور که از فرمول برمیاید رابطه‌ای منفی بین ضریب اطمینان و SEM وجود دارد. هر چقدر این ضریب بالاتر باشد، SEM کوچکتر خواهد بود. مثلاً، اگر ضریب اطمینان کامل باشد، $r = 1$ ارزش عددی SEM برابر با صفر خواهد بود بدلیل اینکه $(1-r)$ برابر با صفر است. بر همین قیاس، هر چقدر ضریب اطمینان کمتر، بهمان اندازه SEM بزرگتر است.

مفهوم SEM در آزمون خیلی مهم است. فرد در استفاده از نمره آزمون خاصی خیلی باید با احتیاط عمل کند، زیرا که بزرگی SEM بر دقت ارزیابی نمره اثر میگذارد. با مثالی این نکته را روشن خواهیم ساخت. تصور کنید که آزمون واژگان در مورد گروهی از دانش‌آموزان به‌اجرا درآمده و نتایج زیر حاصل شده

$$X=20 \quad S_x=5 \quad r=0.84$$

SEM به این ترتیب قابل محاسبه می‌باشد:

$$\begin{aligned} SEM &= S_x \sqrt{1-r} \\ &= 5 \sqrt{1-0.84} \\ &= 5 \sqrt{0.16} \\ &= 5(0.04) \\ SEM &= 2 \end{aligned}$$

این بدان معنی است که خطای معمول در نمره دادن به آزمون واژگان برابر با عدد ۲ است. بدین ترتیب وقتی که فرد به بررسی نمره معینی مینشیند، باید دقت داشته باشد که میانگین نمره ظاهری ممکن است پائین‌تر یا بالاتر از نمره حقیقی آزمون شونده باشد. میزان بالا بودن یا پائین بودن میتواند از طریق ارزش SEM پیش‌بینی شود. معمولاً، یک برآورد درست آنست که نمره واقعی را با توجه به متغیر ۱+ SEM-1 از کل نمره ظاهری ارزیابی کنیم. طبق فرمول زیر:

$$SEM \pm \text{نمره ظاهری} = \text{یک نمره خیلی دقیقتر}$$

چنانچه نمره آزمون فردی مشخص روی برگه آزمون مثلاً ۲۵ باشد، نمره حقیقی وی باید بین ۲۳ و ۲۷ در نوسان باشد.

$$\begin{aligned} SEM \pm 1 &= 25 \pm 2 \\ &= 25 + 2 = 27 \\ &= 25 - 2 = 23 \end{aligned}$$

بنابراین نمره ظاهری نمیتواند بمنزله دقیق‌ترین برآورد توانائی فرد محسوب گردد. بزرگی SEM بایستی در بررسی تمامی نمرات آزمون مورد توجه قرار گیرد. چرا که ارزش SEM از طریق ارزش ضریب اطمینان محاسبه میشود. برای محاسبه SEM فرد بایست در ابتدا ضریب اطمینان را محاسبه نماید. در هر حال کار چندان ساده‌ای نیست که بشود ضریب اطمینان مجموعه‌ای از نمرات یک آزمون

را معین کرد. در واقع، تقریباً محاسبه بزرگنمایی واقعی ضریب اطمینان بدلیل غیرممکن بودن نمره حقیقی در محاسبه امری ناممکن است. این البته بمعنی آن نیست که برای نزدیک شدن به نمره حقیقی دست از تلاش باید برداشت. برعکس، بایست بیشترین تلاش را کرد تا برآورد درستی از نمره واقعی بر مبنای نمره ظاهری بعمل آرند. مضمون این بیان اینست که ارزش ریاضی ضریب اطمینان همیشه بجای اینکه محاسبه شود تخمین زده میشود، سوای دقت نظری، شیوه‌های عملی مختلفی برای برآورد ضریب اطمینان وجود دارد. این شیوه‌ها در زیر تشریح خواهد شد.

۳-۶ شیوه‌های برآورد ضریب پایایی یک آزمون

پایایی یا وثوق بمعنای ثبات نمرات یک آزمون می‌باشد. اصطلاح ثبات نیازمند شفاف‌سازی بیشتری است. تصور کنید که آزمونی در یک روز خاص برگزار شده باشد. چطور میشود مطمئن بود که نمرات آزمونی اگر مجدداً برای همان آزمون شوندگان برگزار شود نمرات ثابتی خواهد بود؟ از توضیحات پیشین معلوم میشود که از آزمون شونده هیچ انتظار نمی‌رود تا پس از برگزاری دو امتحان یکسان دقیقاً همان نمره را دریافت کند. تفاوت ما بین نمرات دو امتحان همانطور که قبلاً تشریح شد، منجر به عدم اطمینان به نمرات و درجه‌ای از خطا در نمره دادن می‌شود.

بعلاوه گفته شد که ضریب اطمینان عبارت است از تمامی واریانس نمرات آزمون منهای واریانس خطا. بدین ترتیب برای ارزیابی ضریب اطمینان یک آزمون بایست واریانس نمرات آزمون را بدست آورد. یکی از راههای بدست آوردن این واریانس استفاده از فرمول زیر می‌باشد.

$$V = \frac{\sum (X - \bar{X})^2}{N - 1}$$

بهر حال این فرمول مقدار واریانس قابل مشاهده را اندازه می‌گیرد نه واریانس واقعی را، بنابراین کافی نیست. راه دیگری برای حساب کردن واریانس از طریق شیوه‌های ارتباطی است. از مبحث ضریب همبستگی در فصل ۴ باید بخاطر آورد که مربع همین ضریب مقدار واریانس معمول بین دو مجموعه نمرات می‌باشد. البته، در ارتباط با همبستگی دو مجموعه نمره از طریق اجرای دو آزمون مختلف درباره دو گروه خاص بدست می‌آید. هنگامیکه همین آزمون دوباره در مورد همان گروه به اجرا درمی‌آید ضریب همبستگی و نه مربع همبستگی همان مقدار واریانس معمول می‌باشد.

واریانس معمول بین دو مجموعه نمره بدست آمده از همان آزمون مقدار ثابت متغیر نمرات آزمون یعنی همان ضریب اطمینان می‌باشد. بنابراین، برای محاسبه ضریب اطمینان نمرات آزمون ضریب همبستگی بین دو مجموعه نمره اکتسابی از دو آزمون همانند در مورد همان دو گروه بایست محاسبه گردد. آزمون‌های دیگری هم به شیوه‌های مختلف امکان دارد. این وجوه ممکن منشاء همان روش‌های مختلف برآورد ضریب اطمینان می‌باشد. آنچه که در زیر آمده شرحی است به اختصار از چهار شیوه

محاسبه ضریب اطمینان. سه تای آنها حاصل استفاده از روش‌های همبستگی است، و یکی دیگر حاصل شیوه غیرهمبسته و فرمولی خاص.

۱-۳-۶ شیوه آزمون و آزمون مجدد

همان‌طور که از عنوان پیداست، ضریب اطمینان بطریق آزمون از گروهی خاص در دو مرحله و محاسبه همبستگی بین دو مجموعه نمره بدست آمده می‌باشد. بهنگام استفاده از این شیوه، فرض بر این است که هیچ تغییر عمده‌ای در معلومات آزمون شونده در فاصله دو امتحان پدید نیاید. از آنجا که بین دو امتحان زمانی منطقی در نظر گرفته میشود، این نوع ضریب اطمینان بنام پایائی یا ثبات نمرات برحسب زمان خوانده میشود. مقدار این پایداری از طریق محاسبه ضریب همبستگی بین دو مجموعه نمره حاصل از برگزاری همان دو امتحان در مورد همان گروه بدست می‌آید. ضریب بستگی همان مقدار ضریب اطمینان است.

گرچه روش آزمون و آزمون مجدد موجبات محاسبه منطقی ضریب اطمینان نمرات آزمون را فراهم می‌سازد، اما معایبی هم دارد. نخست اینکه به دو مرحله آزمون نیازمند است. بدیهی است که برگزاری یک امتحان دومرحله‌ای در یک مرحله برای همان یک گروه عملاً غیرممکن می‌باشد. علاوه، آماده‌سازی شرایط همسان برای برگزاری امتحان در دو مرحله از مسائلی است که بر مشکلات این روش می‌افزاید. تنها در شرایط استثنائی است که میشود دو مرحله امتحان مشابه را برای همان یک گروه برگزار کرد.

دوم، اینکه افراد انسانی موجوداتی متفکر و اهل عملند، آنها همیشه در فرایند یادگیری و آموزش فعالند. بدین‌گونه توانمندیهای آنها اکثراً از آزمونی به آزمون دیگر، بخصوص زمانی که فاصله بین دو مرحله امتحان طولانی‌تر است، در حال تغییر و تحول است. برای اجتناب از تحولات جهش بایستی فاصله امتحانات کوتاه زمان‌بندی شود. در هر صورت، یک ارتباط مستقیم وجود دارد هر چقدر این فاصله بلندتر باشد بهمان اندازه رفتار آزمون شونده متحول‌تر خواهد شد. اما حافظه بعنوان یک فاکتور دخالت کمتری خواهد داشت. این فاصله هر چه کوتاه‌تر، چندان تغییری در رفتار آزمون شونده مشاهده نخواهد شد. اما حافظه بیشتری درگیر خواهد بود. برای حفظ تعادل، استادان فاصله زمان دو هفته‌ای را مناسب میدانند.

سوم، مسئله اگر آزمون قبلی است که ممکن است به‌جا مانده باشد. بخصوص وقتی که فاصله زمانی کوتاه است. یعنی اینکه آزمون شونده ممکن است بگونه‌ای متفاوت در امتحان دوم عمل کند یا مطالبی را از آزمون اولی آموخته است یا اینکه حافظه‌اش هنوز چیزهایی را بیاد دارد. برای رفع این معایب مربیان کوشیده‌اند تا روش محاسبه ضریب اطمینان را بنام روش اشکال همسو بعنوان راه چاره ابداع کنند.

۲-۳-۶ شیوه اشکال همسو

برای اجتناب از مشکلات روش آزمون و آزمون مجدد، مربیان شیوه اشکال همسو را ابداع کرده‌اند. معایب عمده روش آزمون و آزمون مجدد همان مشکل اجرای آزمونی واحد در مورد همان یک گروه در دو مرحله می‌باشد.

در شیوه اشکال همسو، بهر حال، دو شکل همسان، یا همسوی یک آزمون واحد در مورد یک گروه از آزمون شوندگان تنها در یک مرحله به‌اجرا درمی‌آید. آنگاه ضریب همبستگی بین نمرات بدست آمده از دو شکل همان برآورد ضریب اطمینان نمره آزمون خواهد بود.

گرچه این روش مشکل دو مرحله‌ای بودن یک آزمون را رفع می‌کند، اما مشکل دیگری را در قبال باعث می‌شود. تنظیم دو شکل همسو (همتراز) یک آزمون واحد کار ساده‌ای نیست. دو شکل همسوی در آزمون بطور قطع پاره‌ای خصوصیات آماری را نشان می‌دهد. این خصوصیات آنقدر فنی هستند که در این کتاب جایی برای بحث پیرامون آن نمی‌باشد در هر حال پاره‌ای ملاحظات منطقی را باید متذکر شد.

نخست، جدول مختصات برای هر دو شکل آزمون یکی خواهد بود. این بدین معنی است که تمامی موضوعاتی که آیتم‌های آزمون براساس آنها تنظیم می‌شوند در هر دو شکل یکسان خواهند بود. مثلاً اگر یک شکل تعداد معینی آیتم که مبحث خاص گرامر را اندازه می‌گیرد داشته باشد، شکل دوم باید با همان آیتمها و همان مباحث گرامری تنظیم گردد.

دوم، مباحث دو آزمون یا موضوعات آن باید همانند باشند. یعنی اینکه اگر یکی دارای مباحث گرامر، واژگان، و درک مطلب است دیگری هم باید چنین باشد و بهمان نسبت. این همسانی بدان مفهوم نیست که شکل ظاهری آیتمها باید همانند باشد. هر آیتمی در یک شکل دارای بنی است که با بن شکل دیگر متفاوت است. آنچه که در تمامی آیتمها ثابت می‌ماند مبحثی است که قرار است آزموده شود. مثلاً آیتم‌های زیر می‌توانند هم جهت باشند زیرا که هر دو توانائی آزمون شونده را برای تشخیص شکل درست زمان حاضر فعل می‌سنجند.

1- Ali usually late at night.

- a) study b) studies c) studying

2- Reza often The shopping in the afternoon.

- a) do b) does c) doing

با وجود آنکه شیوه اشکال همتراز اصلاحی بر روش آزمون پشت آزمون است بازهم دارای معایبی است. بنابراین، شیوه دیگری بنام دو نیمی یا split-half ابداع شده است.

۳-۳-۶ شیوه دو نیمی - Split-half

اشکال عمده روش آزمون پشت آزمون این بود که یک آزمون واحد در مورد یک گروه در دو مرحله به اجرا درمی‌آید. و عیب اساسی شکل همتراز (همجهت) این بود که یک آزمون با دو شکل همسو به

اجرا درمیآید. همانطور که قبلاً مورد بحث قرار گرفته این دو مشکل همراه عوامل دیگر موجب شده تا محاسبه ضریب اطمینان بدین طریق مشکل شود. بنابراین استادان شیوه Split-half یا دو نیمی را بعنوان یک جایگزین مناسب ابداع کرده‌اند.

اگر راستش را خواسته باشید روش Split-half براساس فرضیه اشکال همتراز پایه‌ریزی شده است: البته، اشکال همتراز آیتماها در یک آزمون واحدند، دو آزمون جداگانه پشتوانه فکری روش دو نیمی یا Split-half این است که آیتماهای متشکله یک آزمون هم‌جنسند. یعنی اینکه تمامی آیتماهای یک تست قرار است تا موضوعات یک مبحث خاص را اندازه بگیرند، مثلاً زمان‌ها، حروف اضافه، نکات دیگر گرامری، واژگان، قرائت و درک مطلب شنیداری که اجزای فرعی مبحثی هستند بنام توانائی زبانی. در واقع این شیوه قبول دارد که یک نوع سنخیت درونی بین آیتماها وجود دارد.

بدین ترتیب رابطه بین آیتماها با توجه به رابطه درونی، نوعی ضریب اطمینان نمرات خواهدبود. بهمین دلیل است که این شیوه گاهی بنام ثبات درونی نمرات آزمون معروف است. در این روش، زمانی که یک آزمون نهائی با آیتماهای یک جنس در اختیار یک گروه آزمون شونده قرار میگیرد، سئوالات تقسیم و از هم جدا میشوند و بدو نیمه درمیآیند. بستگی بین دو نیمه محاسبه ضریب نمره آزمون است.

در استفاده از این روش، دو نکته مهم باید مورد توجه قرار گیرد. اول، روش تقسیم آزمون به دو نیمه و دوم محاسبه ضریب اطمینان کل آزمون از طریق ضریب اطمینان هر نیمه اجازه بدهد تا مورد اول را مطرح سازیم. در یک آزمون با طراحی سنجیده توالی آیتماها از آسان به مشکل است. پاسخ حاضر و آماده به سئوال «چه آسان است؟» یا «مشکل کدام است؟» وجود ندارد. اما در هر حال راز و رمز فرایند روانی بهنگام پاسخ دادن به یک آیتما میتواند بعنوان یک معیار قابل قبول بکمک آید. بدین ترتیب آیتماها میتوانند از نوع تشخیصی تا بیانی مرتب شوند.

با قبول اینکه در آزمون‌های یکصد آیتمی ترتیب آیتماها از ساده به پیشرفته امری مشکل‌تر شده است، تقسیم آزمون به دو نیمه، یعنی از ۱ تا ۵۰ و از ۵۱ تا ۱۰۰ چندان صلاح نخواهدبود. در چنین مواردی، نیمه اول از نیمه دوم ساده‌تر خواهدبود. بدین ترتیب برگزاری امتحان برای هر دو نیمه بهم ارتباطی نخواهدداشت. علاوه بر این، کاملاً امکان دارد که تمامی آزمون‌دهندگان فرصت کافی پیدا نکنند که به همه سئوالات پاسخ گویند و در نتیجه جاهائی خالی بماند و منجر به نابرابری در دو نیمه گردد. برای اجتناب از مشکلات فوق باید روش مناسبی برای تقسیم آزمون به دو نیمه مساوی طرح شود. این امر با انتخاب اعداد فرد برای نیمه اول و اعداد زوج برای نیمه دیگر میسر می‌شود. از طریق این روش، سوال‌های آسان و مشکل بطور مساوی در دو نیمه توزیع می‌شوند. همچنین اگر سوال‌های آخر بی‌پاسخ بمانند، آنها نیز بطور مساوی بین دو نیمه توزیع می‌شوند.

نکته دوم، با توجه به اینکه در روش دو نیمی یا Split-half مسئله جبران افت ضریب اطمینان بدلیل همین دو نیمه شدن آزمون است، از آنجا که بلندی آزمون، تعداد آیتماها، عامل مهمی در ضریب

اطمینان نمره آزمون می‌باشد، تقسیم آزمون به دو نیمه، طول آزمون را به نصف تعلیق خواهد داد. بدین ترتیب، همبستگی بین دو نیمه آزمون، عبارت از ضریب اطمینان یک نیمه خواهد بود، نه تمامی آزمون. پس برای محاسبه ضریب تمامی آزمون از فرمول زیر بنام فرمول Spearman استفاده میشود.

$$r(\text{total}) = \frac{2(0/80)}{1 + 0/80} = \frac{1/60}{1/80} = 0/88 \quad r(\text{total}) = \frac{2(r\text{half})}{1 + (r\text{half})}$$

چنانچه، برای نمونه، ضریب اطمینان نیمی از آزمون ۰/۸۰ حساب شود، ضریب تمامی آزمون چنین خواهد بود.

برای اجتناب از اشکالات فوق، یک روش مناسب برای تقسیم آزمون به دو نیمه باید ابداع میگشت. این کار با انتخاب آیتمهای ساخت برای یک نیمه و آیتمهای آسان برای نیمه دیگر امکان می‌یابد. از طریق این روش، آیتمهای آسان و مشکل بطور یکسان در دو نیمه توزیع خواهد شد. گذشته از این اگر آیتمهای آخری سفید بمانند، باز میتوانند بطور برابر در دو نیمه توزیع گردند. دیگر بلحاظ منطقی باید روشن شده باشد که ضریب اطمینان آزمون کل بالاتر از ضریب نیمه آزمون است.

شیوه دو نیمه محاسبه ضریب اطمینان دارای پاره‌ای مزیت‌ها بر سایر روش‌هاست. نخست، اینکه از دیگر روش‌ها عملی‌ترست. در استفاده از شیوه دو نیمه نیازی به اجرای آزمون در دو مرحله نیست. ضرورتی هم به تهیه دو شکل همسو برای همان آزمون نیست. برای استفاده از روش دو نیمه برگزاری یک مرحله‌ای آزمون کفایت می‌کند.

بهر حال همراه با این محاسن به شیوه دو نیمه دارای پاره‌ای معایب هم هست. اساسی‌ترین نقطه ضعف این روش تألیف آزمون با آیتمهای همسان می‌باشد، زیرا که فرض برابری بین دو نیمه همیشه فرض درستی نیست. بعلاوه، مباحث فرعی، در آزمون مثل گرامر، واژگان، قرائت یا درک مطلب شنیداری، یکدستی آزمون را دچار اشکال می‌سازد. و بدین ترتیب ضریب اطمینان نمره آزمون را کاهش میدهد. برای رفع این اشکال کارشناسان آمار شیوه مناسب‌تری ابداع نموده‌اند.

۴-۳-۶ شیوه KR-21

کودار و ریچاردسون^۱، دو کارشناس آمار معروف یک مجموعه فرمول ریاضی برای محاسبه آماری ابداع کرده‌اند. یکی از این فرمول‌ها، بنام KR-21 برای برآورد ضریب اطمینان نمرات آزمون می‌باشد. این فرمول بر این فرض استوار است که تمامی آیتمها در یک آزمون برای این تنظیم شده‌اند مبحث واحدی را بسنجند. این شیوه، گاهی بنام شیوه تعادل منطقی خوانده میشود، یک شیوه آماری محض

فقط نیاز به عدد میانگین ($\bar{X}$) و واریانس (V) نمرات آزمون دارد. این پارامترها در فرمول زیر قرار داده می‌شوند.

$$(KR - 21)r = \left[\frac{K}{K-1} \right] \left[1 - \frac{\bar{X}(K - \bar{X})}{KV} \right]$$

در حالیکه

K = the number of the items in a test. تعداد آیتمهای آزمون

$\bar{X}$ = the mean score نمره میانگین

V = the variance واریانس

این روش بلحاظ آنکه فاقد هرگونه مشکل و پیچیدگی است بر تمامی روش‌های دیگر برتری دارد. این نیاز به اجرای دو مرحله‌ای مثل آزمون پشت آزمون، اشکال همسو یا شیوه دو نیمه‌ای ندارد. بعلاوه تمامی شیوه‌های دیگر نیاز به استفاده از ضریب بستگی دارد. البته همه مربیان و معلمان با این تکنیک آماری آشنائی ندارند. بنابراین KR-21 عملی‌ترین، متداول‌ترین و مناسبترین روش برای محاسبه ضریب همبستگی است. مثلاً، اگر یک آزمون ۱۰۰ آیتمی در مورد یک گروه از آزمون‌دهندگان به‌اجرا درآید و منتهی به میانگین ۶۰ و واریانس ۴۸ شود، ضریب اطمینان این آزمون میتواند بسادگی بقرار زیر باشد. باید تذکر داد زمانی‌که برای ارزیابی ضریب اطمینان یک سری نمرات روش‌های متفاوتی وجود دارد. سؤال اینست که کی و چرا یک روش خاص بایست مورد استفاده قرار گیرد. برای اینکه روش را بر روش دیگر برتری دهید، کارکرد هر روش بایست مورد توجه قرار گیرد. مثلاً، روش امتحان پشت سرهم مناسبترین است وقتی‌که ثبات نمرات طی یک فاصله از اهمیت برخوردار است. بعبارت دیگر، این‌طور گفته میشود که روش امتحان پشت سرهم ثبات نمرات را برحسب فاصله زمانی نشان میدهد. برعکس روش اشکال همتراز زمانی مفید است که ثبات نمرات برحسب اشکال مختلف از هم مهمتر است. و بالاخره، وقتی پاسخ به تمامی آیتمها همراه هم پراهمیت است، پایداری و ثبات درونی، روش دو نیمه‌ای، KR-21، یا گونه‌های دیگر این فرمول از مناسبترین است.

درک کارکرد هر روشی البته کار انتخاب آن روش را از میان دیگر روش‌ها آسان می‌سازد. در هر حال در هر مورد، محاسن و معایب هر روشی بایست مورد توجه قرار داد. پاره‌ای عوامل منطقی نظیر برگزاری جلسات سخنرانی بر انتخاب یک روش خاص تأثیر خواهدداشت، در هر حال از نقطه نظر عملی، KR-21 همانطور که گفته شد عملی‌ترین و مناسبترین روش محاسبه ضریب اطمینان نمرات می‌باشد.

۴-۶ عوامل مؤثر بر پایایی

همانطور که قبلاً شرح داده شد، ضریب اطمینان، عددی است برای برآورد حد ثبات نمرات، این ثبات برحسب زمان است. امتحان پشت سرهم یا دو شکلی یا اشکال همتراز یا دو نیمی و روش KR-21

. در هر یک از موارد برای محاسبه ضریب اطمینان، یک یا دو مجموعه نمرات از هر گروه آزمون شونده باید در دست باشد. بدین ترتیب، دو فاکتور منجر به ضریب اطمینان آزمون میشود:

آزمون و آزمون شونده. نوسان هر یک از این عوامل بنوعی میتواند بر بزرگنمایی ضریب اطمینان مؤثر باشند. بنابراین این عوامل بطور خلاصه مورد بحث و بررسی قرار خواهند گرفت زیرا که داشتن اطلاعاتی درباره هر یک از این عوامل به آزمون سازان کمک می‌کنند تا در کنترل این عوامل دقیقتر عمل کنند.

۱-۴-۶ تأثیرگذاری آزمون دهنده

فرایند اندازه‌گیری یک خصیصه (موصوف) از یک حوزه علمی به یک حوزه دیگر فرق میکند. در علوم فیزیک مثلاً، موصوف‌ها اغلب اشیاء یا مختصات مربوط به آنها می‌باشند که معمولاً بر اثر پاره‌ای روش‌های کنترلی و دخالت انسان دچار تغییر و تحول میگردند. بنابراین، نوسان در مختصات اشیاء میتواند بطور عینی و مستمر مورد ارزیابی قرار گیرد.

در علوم انسانی، بهر حال موصوفی (ویژگی) که قرار است سنجیده شود دارای خصوصیات ثابتی نیست. زیرا انسان‌ها موجوداتی فعال و پویاند، آنچه که هم به انسان استناد می‌شود نیز از همین پویایی و فعال بودن نصیب برده است. این بدان معنی است که کار و فعالیت انسان بلحاظ ماهیت آن، از زمانی به زمان دیگر یا از مکانی به مکان دیگر فرق میکند. نوسانات در شرایط روانی و فیزیولوژیک آزمون شونده به کار و فعالیت آنان اثر میگذارد. این تأثیرات ممکن است نمره واقعی یا نمره‌ی خطای آنها را افزایش دهد. در مورد قبلی، ضریب اطمینان ممکن است زیاد برآورد گردد و در مورد بعدی ممکن است کم حساب شود.

یک عامل دیگر در ارتباط با آزمون شونده که بر ضریب اطمینان مؤثر واقع میشود همان بودن توانائی آزمون شونده از لحاظ موصوف سنجش شده است. همانطور که قبلاً گفته شد، ضریب اطمینان کارکرد مقدار واریاسیون کار و فعالیت آزمون شونده است. اگر آزمون شوندگان بلحاظ توانائی که قرار است سنجیده شود همسان باشند، نمرات آنها تغییرات چندانی را نشان نخواهد داد. یعنی اینکه، ضریب اطمینان کم برآورد میشود. از طرف دیگر، توانائی آزمون شونده در زمینه خاصی که آزموده می‌شود اگر بشدت متغیر باشد، مقدار واریانس افزایش خواهد یافت و بدین ترتیب ضریب اطمینان بالا خواهد بود، بنابراین بهنگام گزینش آزمون شونده بایست دقت مبذول داشت تا یک سری توانائی منطقی در فرد مورد ارزیابی قرار گیرد. والا، دست یافتن به یک برآورد دقیق از ضریب اطمینان کاملاً مشکل خواهد بود.

۲-۴-۶ تأثیرات عوامل آزمون

نوسانات در شرایط روانی و فیزیولوژیک فرد گفته میشود که تأثیری چندان بر ضریب اطمینان ندارد. در این زمینه، کسی ممکن است بگوید که در میان گروه آزمون شونده‌ها افرادی باشند که بطور اتفاقی نمرات واقعی بالائی بگیرند و در مقابل کسانی هم نمرات اشتباه بالائی. این نوسانات اتفاقی یا ممکن است اثر یکدیگر را خنثی کنند یا آنقدر قابل ملاحظه نباشند که بر ضریب اطمینان بطور قاطع اثر بگذارند.

ممکن است عوامل آزمون، از طرف دیگر بمیزان خیلی زیاد و مؤثری بر ضریب اطمینان اثر بگذارند. نوسانات عوامل آزمون ممکن است بدلیل ساختار موضوعی آزمون، روش‌های اجرایی، یا فرایند نمره‌گذاری باشد. هر یک باختصار شرح داده خواهد شد.

۱-۲-۴ ساخت آزمون

یکی از دلایل عدم ثبات نمرات امتحانی مربوط به ساختمان آزمون می‌باشد. بعضی پارامترهای درونی خود آزمون میتواند منجر به حس عدم اطمینان به آن باشد. ممکن است این پارامترها شامل همسانی آیتمها، شتاب در اجرای آزمون یا طولانی بودن آن باشد. البته با قدری دقت و برنامه‌ریزی، اثر این عوامل گرچه زائل نمیشود، اما همواره کاهش می‌یابد.

نخستین عامل همانندی خود آیتمها است. بعنوان یک قانون کلی، آیتمهای یک آزمون باید یک مبحث یا وجوه یک مبحث را بسنجند. مثلاً، آزمون گرامر می‌باید تنها براساس آیتمهای مباحث گرامر تنظیم یابند. اگر پاره‌ای آیتمها بعنوان مثال واژگان را مورد سؤال قرار دهند. آزمون یکدستی خود را از دست خواهد داد. باور عمومی بر این است که هر چقدر آیتمهای آزمون یکدست‌تر باشند بهمان اندازه نمرات ثبات بیشتری خواهند داشت. این باور بدلیل آنکه تقریباً تمامی ضرایب اطمینان و ثبات بروش همبستگی بدست می‌آید، قابل توجیه است. آیتمهای یکدست در راستای هم هستند و بدین ترتیب دارای بستگی بالائی هستند. هر چقدر بستگی بین دو قسمت یک آزمون یا بین آیتمهای یک آزمون بالاتر باشد، ضریب اطمینان آن بهمان اندازه افزون خواهد بود. از سوی دیگر وقتی آیتمهای یک آزمون در تضاد باهم هستند، بستگی بین آنها کاهش و در نتیجه ضریب بستگی پائین است.

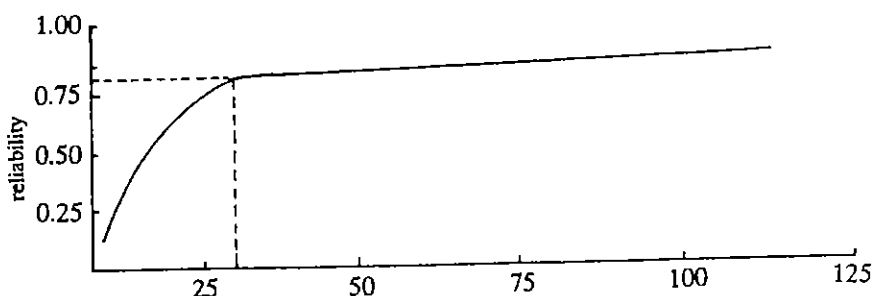
پارامتر دوم شتاب زمانی است که یک آزمون طی آن برگزار میگردد. در آزمون‌های سرعتی، آزمون شونده در مقابل زمان تعیین شده بر قایت برمی‌خیزد. در حالیکه در آزمون‌های قدرتی، آزمون شونده بر محتوا و موضوع آزمون تمرکز دارد. به دیگر سخن، در آزمون شتابی آزمون شونده در چهارچوب یک زمان محدود می‌باید مبادرت به پاسخگویی به سؤالات نماید. حتی اگر جواب همه سؤالات را هم بداند باز ممکن است که نتواند به همه آنها برسد. چرا که وقت کم می‌آورد. در آزمون قدرتی از سوی دیگر عامل زمان در کار نیست. آزمون شونده فرصت دارد تا تمامی آیتمها را پاسخ گوید البته، خیلی مشکل

خواهد بود که امتحانی تحت شرایط محدودیت زمانی یا بخاطر تشخیص توانایی برگزار گردد. اکثر آزمونها در نقطه‌ای بین این دو حد فاصل فرود می‌ایند.

به میزانی که آزمون شتاب‌دار است نمرات اشتباه هم در کار آزمون شونده وارد می‌شود. زیرا که محدودیت زمان موجب می‌شود تا آزمون شونده دچار شتابزدگی گردد. در نتیجه اکثراً تا لحظه آخر در شرایطی آرام و فارغ از دلهره بسر نمی‌برند. البته، اختصاص زمانی نامحدود نیز بر نتایج آزمون اثر منفی خواهد گذاشت. زیرا که آزمون شونده بعوض تمرکز، وقت‌گذرانی خواهد کرد. بنابراین، باید بین تعداد آیتمها و زمان اختصاص یافته توزنی برقرار باشد. بعنوان یک قاعده، برای یک آیتم چند گزینه‌ای ۳۰ ثانیه زمان لازم است. البته در آزمون‌هایی نظیر درک مطلب خواندن، Cloze یا فعالیت‌هایی نظیر آن، نظر به اینکه آزمون شونده بایستی متنی را مثل یک آیتم، بخواند پس وقت لازم یک دقیقه کافی است. پارامتر سؤم عبارت است از تعداد آیتمها یا طول آزمون. آزمون هر چه طولانی‌تر تعداد آیتمها هر چقدر بیشتر، بهمان اندازه آزمون قابل‌اعتمادتر است، توجیه این حس اعتماد و اطمینان همان واریانس نمرات و کار آزمون شونده می‌باشد.

وقتی تعداد زیادی آیتم در یک آزمون وجود داشته باشد آزمون شندگان بطور حتم در مورد این آیتمها متفاوت عمل خواهند کرد و بهنگام نمره گرفتن شدت باهم فرق خواهند داشت در نتیجه، واریاسیونی که نشانه ضریب اطمینان به آزمون می‌باشد، افزایش خواهد یافت.

به این نکته باید اشاره کرد که طول آزمون ضریب اطمینان آنرا تا نقطه‌ای معین بالا خواهد برد. بالاتر از حد معینی تاثیر طول بر ضریب اطمینان بقدر ناچیزی کاهش می‌یابد. ارتباط بین طول آزمون و ضریب اطمینان توسط بسیاری از استادان معلوم شده است. نمودار زیر نشان‌دهنده ارتباط تقریباً نزدیک بین طول و ضریب اطمینان می‌باشد.



گرچه این شکل نموداری تقریبی است، میزان اثر تعداد آیتمها را بر ضریب اطمینان نشان می‌دهد. همانطور که از آن برمی‌آید، بالای ۳۰ تا ۳۵ آیتم، ضریب اطمینان شدت افزایش می‌یابد. بالاتر از ۳۵ بهر حال افزایش در ضریب آنقدر آرام است که قابل چشم‌پوشی است. بعبارت دیگر اگر آزمونی ضریب اطمینانی مثلاً برابر با ۰/۸۰، حدود ۳۰ آیتم نشان دهد. و ضریب اطمینانی برابر با ۰/۸۵ با ۱۰۰ آیتم

خیلی منطقی‌تر است تا از آن ۰/۵ افزایش در ضریب اطمینان چشم‌پوشی کنیم و بدین ترتیب در زمان و نمره دادن تا میزان زیادی صرفه‌جویی نمائیم.

بمنظور محاسبه اثر تعداد آیتمها بر ضریب اطمینان بطریق محاسبه ریاضی، فرمول اسپیرمان براون^۱ بکار می‌آید.

$$r_k = \frac{K_{r1}}{1 + (k-1)r1}$$

در این فرمول، r_k دلالت بر آزمون میکند، هنگامیکه K بارها به نسبت طول اصلی بلندتر میشود $r1$ به پایایی ظاهری آزمون اشاره دارد. باطول فعلی آن K دلالت دارد بر تعداد دفعاتی که طول قرار است افزایش یابد. یکی دو مثال به روشن شدن موارد استفاده این فرمول کمک خواهد کرد.

همانطور که در روش دو نیمی محاسبه پایایی اشاره شد، تصور کنید که آزمون به دو نیمه تقسیم شده و پایایی آن ($r1 = 0/80$) شده باشد. اگر پایایی آزمون باید با طول کامل آن منطبق شود (یعنی دو برابر طول فعلی) یعنی $K = 2$ باشد، پایایی انطباق یافته میتواند طبق فرمول چنین باشد.

$$\begin{aligned} r_k &= \frac{K_{r1}}{1 + (k-1)r1} \\ &= \frac{2(0.80)}{1 + (2-1)0.80} \\ &= \frac{1/60}{1/80} = 0/88 \end{aligned}$$

بنابراین پایایی (ضریب اطمینان) تمامی آزمون، وقتی که تعداد آیتمها دو برابر گشته‌اند، از ۰/۸۰ به ۰/۸۸ افزایش می‌آید.

مثال دیگر، تصور کنید یک آزمون ۴۰ آیتی دارای ضریب اطمینانی برابر با ۰/۷۰ یعنی $r1 = 0/70$ باشد. فرض کنید که ده آیتم به این آزمون اضافه شده باشد. طول آزمون به یک چهارم طول فعلی افزایش می‌یابد ($k = 1 + \frac{1}{4} = 1/25$). ضریب اطمینان آزمون جدید بشرح زیر محاسبه خواهد شد.

$$\begin{aligned} r_k &= \frac{K_{r1}}{1 + (k-1)r1} \\ &= \frac{1/25(0.70)}{1 + (1/25-1)0.70} \\ &= \frac{0.875}{1.175} = 0.74 \end{aligned}$$

^۱ - Spearman Brown

از این رو، اگر طول آزمون را به یک چهارم طول کل آن افزایش دهیم، پایایی (ضریب اطمینان) آن از عدد ۰/۷۰ به ۰/۷۴ افزایش خواهد یافت. در استفاده از این فرمول دو نکته باید مورد توجه قرار گیرد. نخست، اضافه شدن آیت‌ها، فرض تبعیت از اصول همسان بودن است. یعنی اینکه آیت‌های اضافی بلحاظ موضوع و شکل کاملاً شبیه همان آیت‌های آزمون است. والا ضریب اطمینان انطباقی درست محاسبه نخواهد شد. دوم، موضوع «K» در فرمول به تعداد آیت‌های اشاره درست نخواهد داشت. بلکه به افزایش نسبی آیت‌ها با توجه به طول فعلی آزمون دلالت خواهد کرد. یعنی اینکه اگر آزمونی ۲۰ آیت دارد، وقتی که شمار آیت‌ها دوبرابر شود. K مساوی عدد ۲ خواهد شد. اما اگر تنها ۵ آیت اضافه شود K برابر با $1 + \frac{1}{4}$ یا ۱/۲۵ خواهد بود.

اشاره کنیم که فرمول تخمینی اِپسیرمان آزمون سازان را قادر می‌سازد تا تعداد آیت‌هایی که برای تحقق سطح مطلوبی از اطمینان لازم می‌آید تعیین کنند. فرمول برای تشخیص عدد k بقرار زیر است.

$$r_k = \frac{K r_1}{1 + (k-1)r_1}$$

$$K r_1 = r_k + r_k k r_1 - r_k r_1$$

در این فرمول، K تعداد باری است که آزمون بلحاظ طول باید افزایش یابد. Rk ضریب دلخواه و r_1 ضریب ظاهری است.

مثلاً، اگر یک آزمون ۴۰ آیتی پایایی برابر ۰/۶۰ داشته باشد. $r_1 = ۰/۶۰$ و آزمون ساز می‌خواهد پایایی را به ۰/۸۰ برساند یعنی $r_k = ۰/۸۰$ تعداد باری که طول باید افزایش یابد بطریق زیر محاسبه خواهد شد.

$$K = \frac{r_k - (1 - r_1)}{r_1(1 - r_k)}$$

$$= \frac{0/80(1 - 0/60)}{0/60(1 - 0/80)}$$

$$= \frac{(0.80)(0.40)}{(0.60)(0.20)}$$

$$= \frac{0/32}{0/12} = 2/6 \quad \text{بالاخره}$$

پس K تقریباً برابر است با ۲/۶. این بدین معنی است که آزمون باید تا ۲/۶ بار برابر طول فعلی افزوده گردد؛ از آنجائیکه ۴۰ آیت در آزمون وجود دارد؛ ۲/۶ بار ۴۰ میشود ۱۰۴، پس تعداد آیت‌های باید برابر ۱۰۴ افزایش یابد. به سخن دیگر، اگر آزمون ساز دوست دارد که آزمون به سطح ضریب اطمینان ۰/۸۰ صعود کند، بهتر است که بعوض ۴۰، ۱۰۴ آیت داشته باشد.

با استفاده از این فرمول، آزمون سازان در شرایط بهتری خواهند بود، برای اینکه درباره طول آزمون و ضریب اطمینان آن تصمیم بگیرند.

۳-۲-۴ اثر عوامل اجرایی

عدم ثبات و پایداری در فرایند اجرا، خطای اندازه‌گیری را افزایش می‌دهد و بدین ترتیب از ضریب اطمینان آزمون میکاهد، میشود اثر اجرا را موقعی بطور جد دریافت که آزمونی در مورد گروههای مختلف در مکانی متفاوت و شرایطی گونه‌گون به اجرا گذاشته شود. البته برگزاری یک آزمون واحد هم تابع یک چنین نوساناتی است.

پاره‌ای نوسانات ناشی از اجرای آزمون بلحاظ ابهام در آموزش زمان اجرا و میزان تعامل بین آزمون شونده و آزمون کننده می‌باشد. بعضی دیگر به پاره‌ای رفتارهای متعارف یا غیرمتعارف در فرایند اجرا برمیگردد. مثلاً آیا زمان باقی مانده بطور مرتب اعلام شود یا نشود. به سئوالاتی که مطرح میشود پاسخ داده شود یا خیر. آیا توضیحات اضافی طی جلسه امتحان ضرورت دارد یا نه همه میتوانند منشاء خطا باشند.

بعنوان مثال، تصور کنید که آزمونی مبنی بر درک شنیداری که ضمن آن آیت‌های توسط یکی از ممتحنین خوانده میشود باشد. وی تا چه حد میتواند به خواندن ادامه دهد؟ چند بار به لحاظ فیزیولوژی گفتار دچار حالت تهق خواهد شد؟ تمامی این مسائل بر نتایج آزمون اثر خواهند گذاشت. تازه یکی دیگر از مسائلی که اختلال ایجاد میکند مسئله مکان اجرای آزمون است. آمد و رفتی که موجب اختلال در نظم و آرامش جلسه و پرت شدن حواس آزمون‌دهندگان میشود یا نامناسب بودن نور، حرارت و از این قبیل میتواند بر حسن انجام امتحان و بدین ترتیب بر ضریب اطمینان آن خدشه وارد سازد. پس بعنوان یک مجری فرد باید سعی کند که اثر فاکتورهای اجرایی را کاهش دهد. شیوه اساسی برای کنترل این عوامل اینست که مراحل اقدامات اجرایی را قبلاً برنامه‌ریزی کنند. چنانچه اصول روشهای اجرایی از پیش بطور دقیق برنامه‌ریزی گردد، بیشتر اختلالات رفع میگردد.

۳-۲-۴ اثر عوامل نمره‌گذار

همانطور که در فصل پیشین گفته شد، یک آزمون میتواند هم بطور عینی و هم نظری نمره‌گذاری شود. در یک آزمون بطور عینی نمره داده شده، خوش آمدن یا بد آمدن نمره‌گذار بر نتایج اثری نخواهد گذاشت. بنابراین، از لحاظ نمره اختلال یا نوسانی در کار نیست. در چنین مواردی، ضریب اطمینان هم تحت تأثیر نمره‌گذاری قرار نمی‌گیرد. در آزمون‌هایی که نمره‌گذاری منشاء نظری دارد. بهر حال، بلحاظ تفاوت و نظریات بهنگام نمره دادن، وجود خطا بر ضریب اطمینان اثر میگذارد مثلاً، اگر قرار باشد متن یک سری انشاء خوانده و نمره داده شود. در چنین صورتی دو عامل خطا در فرایند نمره دادن ممکن است وجود داشته باشد. نخست اشتباه ناشی از نوسان نمره بهنگام دوبار تصحیح

کردن انشاء. بدیهی است که یک فرد در دو بار و شرایط متفاوت نمیتواند یک نمره یکسان به انشاءای بدهد. تفاوت ناشی از این نوسان ریشه در یک حال باقی نماندن ممتحن دارد. به کلام دیگر، این خطا از درون وی سرچشمه میگردد. بهمین خاطر درون‌زا خوانده میشوند. دوم خطای ناشی از حالات مختلف دو یا چند نمره دهنده متفاوت است. یعنی باز دو نمره گذار یک نمره واحد نخواهند داد. یک چنین خطایی که بدلیل تفاوت‌های بین دو نمره می‌باشد، خطای بین‌الطرفین است خطای درون فردی یا خطای بین فردی واریانس اشتباه را در محاسبه ضریب اطمینان افزایش میدهد.

البته، رعایت پاره‌ای توصیه‌ها ممکن است به کاهش اثر بعضی روش‌های نمره‌گذاری کمک کند. برای آنکه این‌گونه خطاهای نمره‌گذاری را به حداقل برسانیم، نمره گذاره باید تشویق به رعایت معیارهای روشن تعریف شده شود که به متن بدون توجه به نویسنده آن، بدون سابقه آشنائی و بدور از هرگونه پیش‌داوری نمره دهد. و بیش از یکبار نمره بگذارد این رهنمون، البته، اثر این گونه نمره‌گذاری را کاهش خواهد داد، گرچه خطای این نوع نمره دادن را بطور کلی از بین نخواهد برد.

۵-۶ اعتبار بخشیدن (اعتبار)

دومین خصوصیت عمده یک آزمون درست و حسابی معتبر بودن آنست. در فصل قبل از مضمون ضریب اطمینان تعریفی بدست داده شد و راه‌هایی برای محاسبه آن ارائه گشت. چنانچه موضوع اعتبار دادن بهمان روشی که گفته شد رعایت گردد نتیجه دلخواه خواهد بود. متأسفانه در مورد اکثر آزمون‌ها چنین نیست. بلحاظ آنکه معتبر بودن بیشتر از ضریب اطمینان موضوعی متکی بر آزمون است. علاوه بر این ضریب اطمینان یک پارامتر کاملاً آماری است. یعنی اینکه، کاملاً جدای از خود آزمون میتواند معین گردد، مثلاً یک سری نمرات شاخص یک آزمون مشخص، درجه این ضریب بدون حتی رجوع به آزمون میتواند تعیین گردد. معتبر بودن، از طرف دیگر غالباً مربوط به خصوصیات آزمون است. بنابراین، تنها از طریق اکتساب چند تا نمره معین از کل یک آزمون شخص را قادر نمیسازد که برای آن اعتباری فراهم سازد. معتبر بودن دلالت بر چهارچوبی میکند که آزمون در آن چهارچوب موضوعی را که میخواهد اندازه بگیرد، میگیرد. در فصول گذشته، موضوع کارکرد آزمون، یعنی منظوری که آزمون برحسب آن تنظیم گردیده تشریح شد، موضوع معتبر بودن، بهمین سان به این موضوع ارتباط دارد که آیا آزمون به هدفی که خواسته رسیده یا خیر. بعنوان مثال، اگر آزمونی در صدد است تا توانائی دانش‌آموزان را بر مبنای مبحثی بخصوص محک زند، مطلوب آنست که ببینیم آن آزمون در عمل اطلاعات لازم پیرامون مبحث موردنظر را فراهم ساخته یا به مطلب دیگری پرداخته است. برای آنکه ببینیم آزمونی دارای اعتبار هست یا نه این کافی نیست که آزمون را از نقطه نظر ریاضی یا آماری مورد توجه قرار دهیم. بلکه سایر جنبه‌های امتحانی هم از اهمیت برخوردارند، این است که انواع مختلف شیوه‌های اعتباری برای بررسی وجوه مختلف آزمون وجود دارد. کلاً گفته باشیم که سه نوع وجه

اعتباری وجود دارد اعتبار موضوعی، اعتبار معیاری و اعتبار ساختاری. هر یک به اختصار در زیر مورد بحث قرار گرفته است.

۱-۵-۶ اعتبار موضوعی

اعتبار موضوعی برمیگردد به انطباق موضوعی آزمون با موضوع مطالب درسی‌ای که قرار است آرموده شود. موضوع مطالب درسی ممکن است آنچنان تمرین گسترده‌ای از آن بدست دهد که هم محتوای مواد درسی را شامل شود و هم اهداف آموزشی را. محتوای مواد درسی با موضوعات یا همان حوزه درسی مربوطند که در آن حوزه مطرحند. اهداف آموزشی با درجه‌ای از یادگیری که دانش‌آموزان قرار به رسیدنش دارند مربوط است. هر دو این وجوه بایست در بررسی معتبر بودن موضوع (محتوا) به حساب آورده شوند. بنابراین اعتبار داشتن موضوع را میشود بعنوان حدی که یک آزمون بمنزله نمونه شاخص محتوایی که در یک سطح معین از یادگیری اندازه گرفته میشود تعریف کرد. پس، نقطه ثقل معتبر بودن مضمون، تناسب نمونه عواملی است که در آزمون وجود دارد. اصطلاح تناسب هم مربوط است به تناسب نمونه سئوالات و میزان یادگیری. این است که معتبر بودن موضوع بعضی وقتها مناسب بودن آزمون بحساب میاید.

برای آنکه آزمونی بلحاظ موضوع معتبر بماند، بایست دو معیار عمده را مدنظر داشته باشد. اول موضوع آزمون بایست متناسب و منطبق با محتوای مطالب درسی که آرموده میشود باشد. دوم هدف از آزمون باید سنجش سطح متناسبی از آموخته‌های دانش‌آموزان باشد. مثالی به روشن شدن این نکته ممکن است کمک کند:

تصور کنید که آزمونی قرار باشد برای اندازه‌گیری توانائی آزمون‌شوندگان مبنی بر تشخیص گرامر درست یا شیوه صحیح زبان به‌اجرا درآید. برای اطمینان از این امر که موضوع آزمون بر مبحثی مثلاً استفاده درست گرامر منطبق است، باید باصطلاح مثال مشتمل نمونه خرواری از آیت‌های گرامری باشد. این از طریق استفاده از جدول مختصاتی که قبلاً در فصل چهار مورد بحث قرار گرفت ممکن میگردد. آزمون نه فقط باید نشانه‌ای از موضوع بلکه نشانه‌ای از اهمیت یا وزن موضوعات پاره‌ای عناصر زبانی نیز باشد. علاوه بر این، آیت‌های آزمون باید متناسب با سطح آموخته‌هایی باشد که آزمون‌دهندگان انتظار دارند که بطور دقیق آرموده شود. آیت‌های سخت و پیچیده متناسب با دانش‌آموزان مقطع راهنمایی نیست.

بدین ترتیب در رابطه با معتبر بودن موضوع، هم موضوع آزمون و هم دانش‌آموزانی که برای ایشان آزمون تنظیم میشود به حساب آورده شوند. اعتبار داشتن موضوع از مهمترین نوع معتبر بودن است که میتواند از طریق بررسی سنجیده موضوع آزمون حاصل آید. گرچه توضیح عددی رایجی دایر بر اعتبار موضوعی وجود ندارد اما مفیدترین اطلاعات نظری درباره مناسب بودن یک آزمون را در اختیار میگذارد.

البته، ذهن‌گرایی بخودی خود یک عقب‌گرد به حساب می‌آید. دو نفری که درک یکسانی از موضوع امتحانی ندارند، داوری متفاوتی ممکن است درباره انطباق آیتمها و موضوع امتحانی داشته باشند، یکی از راههای اجتناب از ذهنی‌گرایی یا نظری بودن بی‌اندازه این است که آزمون توسط چند نفر مورد بازبینی قرار گیرد.

راه دیگر تشریح موضوع آزمون به تفصیل تا جایی که ممکن است و انتقال این شرح و تفصیل در یک جدول مختصات می‌باشد. در اینجا باید بین معتبر بودن موضوع که به متناسب بودن موضوع آزمون مربوط است و اعتبار ظاهری گفته شده تفاوت قائل شد، زیرا غالب مواقع این دو نوع معتبر بودن باهم آمیخته میشوند. اعتبار بظاهر هیچگاه بمعنی واقعی کلمه اعتبار نیست. بدیهی است معتبر بودن ظاهری بدین معنی است که آیا آزمون در صورت ظاهر معتبر مینماید یا خیر. یعنی اینکه، آیا برای یک فرد آموزش ندیده در یک نگاه به آزمون امکان دارد که بفهمد که آزمون چه چیزی را میخواهد بسنجد یا نه؟ اعتبار داشتن تظاهر اغلب یک خصوصیت دلخواه برای یک آزمون از نقطه نظر افکار عمومی بمعنی همان مفید بودن است.

البته، اگر آزمون بی‌ربط بنماید که آزمون‌دهنده آنرا جدی نخواهد گرفت. یا آندسته از دانش‌آموزان زرنگ آنرا مفید نخواهند یافت. اعتبار موضوعی، از طرف دیگر، بطور مستقیم به موضوع آیتمهای متشکله یک آزمون مربوط میشود. مثالی این نکته را روشن خواهد ساخت یک آزمون ۲۰ آیتی را در نظر آرید که بر مبنای حروف اضافه در حال تجدیدنظر باشد. مسئله اعتبار ظاهری این خواهد بود که آیا این آزمون یک آزمون گرامری حروف اضافه هست یا نیست. در این مورد البته، پاسخ مثبت است. بدین معنی که آزمون دارای وجهی ظاهری است. بهر حال آیا این آزمون دارای وجه موضوعی هم هست یا خیر کل بستگی به بررسی دقیق انطباق بین آیتمها و جدول مختصات دارد. در صورتی که این آزمون برای اندازه‌گیری کار دانش‌آموزان پس از یک دوره کوتاه آموزش حروف اضافه تنظیم شده باشد، دارای اعتبار موضوعی هم خواهد بود. در غیر این صورت جای سؤال است.

باید اشاره کرد که یک آزمون میتواند تا حدی دارای وجهی موضوعی دلخواه هم باشد، اما البته نه وجهی ظاهری. مثلاً: آزمونی همچون تست کلوز Cloze که به تجربه ثابت شده دارای وجهی منطقی نیز می‌باشد، حداقل به لحاظ شکل ظاهری آزمون خوبی برای مباحث گرامر مثلاً نیست. بدین ترتیب آزمون‌سازان چندان نباید دلمشغول وجهی ظاهری بمانند. اما بهر حال بایست در دادن وجهی موضوعی به یک آزمون در هر حال دقت را رعایت کنند.

همانطور که قبلاً گفته شد، حتی اگر آزمونی ثابت شود که دارای وجهی موضوعی قابل قبولی هم هست، نقد و ارزیابی پیرامون آن اساساً امری نظری است. صرف‌نظر از اینکه چقدر این اطلاعات ممکن است مفید واقع شوند، مربیان میل دارند تا این اطلاعات در قالب اعداد و ارقام درآیند. معهذاً انواع دیگری از این معتبر بودن، نظیر معتبر بودن به معیار و وجهی ساختاری ابداع شده‌اند تا یک چنین اطلاعاتی بدست آید.

۲-۵-۶ معتبر بودن بر پایه معیار

از طریق ابداع وجهی موضوعی، مطمئن خواهیم بود که آزمون از طریق موضوع در خدمت هدف آن است. یعنی اینکه انطباق بین موضوع امتحان و محتوای مطالب درسی وجود دارد. این مطلوب‌ترین خصوصیت یک آزمون است بدون آنکه با چیزی خارج از آن قیاس شود. بنابراین به اتکاء همین معیار از طرفی، به بررسی وجه انطباق بین نمرات اکتسابی از آزمونی تازه و نمرات حاصل از محک و معیاری غیر از آن می‌پردازد. این معیار میتواند از سنجش نظری معلم تا نمره دادن عینی در یک آزمون استاندارد متغیر باشد.

برای رسیدن به اعتبار برپایه معیار، آزمون جدید همراه با در نظر گرفتن معیارهای گفته شده دوباره گروه به‌اجرا گذاشته میشود. آنگاه انطباق یا بستگی بین دو مجموعه نمره حاکی از اعتبار بر پایه‌ی معیار امتحانی می‌باشد. برحسب زمان اجرای آزمون دو نوع معتبر بودن بر پایه معیار ابداع شده است.

۱-۲-۵-۶ اعتبار حاصل از همزمانی امتحانی

این اعتبار نوعی معتبر بودن به اتکای معیار می‌باشد برای رسیدن به اعتبار همزمان آزمونی تنظیم شده تا مباحث خاص را بسنجد و به‌همراه آزمون معتبر شناخته شده دیگری به‌اجرا گذاشته میشود. دو مجموعه نمره، که از آزمون جدید حاصل آمده، باهم همبسته‌اند. درجه همبستگی نشانه اعتبار همزمان آزمون است.

بعنوان مثال، یک آزمون تخصص زبان جدیدالابداع در مقابل آزمونی همچون توفی که قبلاً آزموده شده است میتواند معتبر باشد. سؤال این است که آیا آزمون جدید که برای سنجش زبان تخصصی برنامه‌ریزی شده در واقع قادر به رقابت با توفی می‌باشد و بنابراین دو آزمون به همان یک گروه در یک زمان داده میشود و ضریب بستگی بین دو سری نمره، نشانه معتبر بودن همزمان آزمون جدید است.

۲-۲-۵-۶ پیش‌بینی معتبر

نوع دیگر معتبر بودن برپایه معیار پیش‌بینی معتبر می‌باشد، که با معتبر بودن همزمان پیوند خیلی نزدیکی دارد. پیوند نزدیک دارد زیرا که، این نیز بر نوع انطباق بین نمرات اکتسابی از آزمون جدید و نمرات آزمون قدیم متکی است. با این وجود با اعتبار همزمان این فرق را دارد که در اینجا دو آزمون میتوانند همزمان نباشند. بلکه گاهی با فاصله برگزار شوند.

مثلاً، تصور کنید که آزمونی بدین‌منظور تنظیم شده باشد که موفقیت داوطلب را در امتحانات ورودی سراسری دانشگاه پیش‌بینی کند. یک گروه از داوطلبان اول آزمون جدید را می‌گذرانند بعد هم آزمون ورودی دانشگاه را. انطباق بین دو نمره بمنزله پیش‌بینی قابل اعتبار آزمون است.

دو استفاده از همزمانی یا پیش‌بینی معتبر، یعنی معتبر بودن معیاری، نکاتی چند می‌باید مورد توجه قرار گیرند. نخست مختصات یک آزمون خوب. یعنی اینکه بلحاظ معتبر و مطمئن بودن دارای

ضریب منطقی‌ای باشد. دوم موضوع سنجش همان موضوعی باشد که آزمون جدید میخواهد آنرا بسنجد.

سوم بهنگام ارزیابی ضریب اعتبار رعایت دقت ضرورت دارد. برای مربیان و مسئولان آزمون مهم است که بخاطر داشته باشند که انطباق بین آزمون جدید و آزمون معیار مسئله‌ای مداری است. مثلاً، آزمون معتبر و جدید A در مقابل آزمون پایه B قرار دارد. اما آزمون B بایستی در برابر آزمون دیگری بنوبه خود آزمون معتبر بحساب آید. مثلاً در مقابل آزمون C و بهمین ترتیب این حلقه ادامه می‌یابد. این بدان‌معنی است که معتبر بودن به معیار بایستی با دقت کامل مورد تفسیر قرار گیرد زیرا که تمامی ضرایب اعتبار که از پی یکدیگر معتبر شناخته میشوند، متکی به همان آزمون اولی هستند. نکته چهارم طبیعت تجربی این معتبر بودن است. از طریق معتبر بودن امتحانی است که نمرات در قالب ارقام پدید می‌آیند. این نوع اعتبار بعنوان اعتبار تجربی هم شناخته میشود. برعکس آنکه اعتبار موضوعی غیرتجربی است. زیرا که پدید آورنده داده‌های عددی نیست، اغلب محققان و مربیان بشدت بر اعتبار تجربی بطور عام و بر اعتبار همزمان بطور خاص تأکید می‌ورزند. این امر بدلیل عینی بودن و عملی بودن است.

۳-۵-۶ معتبر بودن ساختاری

انواع مختلف معتبر بودن که بدین ترتیب تا حال شرح داده شد به نوعی به کاربرد ویژه نتایج آزمون در عمل مربوط میشود. اینها اطلاعاتی را در اختیار می‌گذارند که مربیان بتوانند براساسشان تصمیم بگیرند که چگونه یک آزمون یک مطلب خاص را بیازماید یا به چه ترتیب نتیجه امتحانی آزمون‌دهنده‌ای را در آینده پیش‌بینی کند، غیر از یک چنین موارد استفاده عملی فی‌الغور خاص در رابطه با تمهیدات امتحانی مسائل دیگری هم هست که میشود مطرح ساخت. یک سؤال مهم که پیرامون اعتبار یک آزمون پرسیده میشود این است که آیا آنچه را که قرار است بسنجد یا خیر. بنابراین، چنانچه هدف سنجش توانایی زبان گفتاری یک دانش‌آموز باشد و آزمون بتواند یک چنین کاری را به‌انجام رساند در واقع اگر این توانایی را بتواند بسنجد، گفته می‌شود که آزمون معتبری است. اما یک مسئله مهم دیگر در ارتباط با اعتبار آزمون اینست که آیا اصلاً چیزی بنام مهارت زبان کلامی وجود دارد یا خیر. به سخن دیگر، آزمونی هدفش سنجش یک مبحث است. این مبحث در این مورد خاص مبحثی روانشناختی است. سؤال هم آنست که واقعیت این مباحث چگونه است؟ بدین ترتیب اعتبار سازی مربوط میشود به طرح و اندازه واقعی وضعیت ساختی این مبحث روانشناختی.

یک مرور سریع بر آزمون‌های موجود روشن می‌سازد که آزمون‌های بی‌شماری تحت نامهای مختلف مورد استفاده قرار می‌گیرد. مشهورترین آنها در این میان تست‌های هوش، اصول منطق انتقادی، درک مطلب خواندنی، مهارت در مطالعه، و تمامی انواع تست‌های استعداد و رفتار می‌باشد. اعتبار ساختی سئوالاتی درباره چگونگی ساخت واقعی این موارد مطرح می‌سازد.

تعیین اعتبار ساختاری تأثیر مهمی بر بسیاری از آزمون‌ها میگذارد. آزمون‌های مختلفی با اسامی گونه‌گون هستند که در واقع همان ساخت روانشناختی را میسنجند. تعیین اعتبار ساختی امری مشکل است زیرا مستلزم استفاده از تحلیل آماری دقیق بنام تحلیل عاملی است. در هر حال، مربیان باید بخاطر داشته باشند که معتبر بودن بلحاظ ساختی پراهمیت‌ترین نوع اعتبار داشتن است که بر سایرین تفوق دارد. دلیل آن هم ساده است. اگر مبحث یا ساخت خاصی وجود نداشته باشد پس تلاش برای سنجیدن هم خود بخود معنی نخواهد یافت.

اهمیت اعتبار ساختی و مشکلات مربوط به تعیین آن منجر به بحث‌های جالبی بین استادان درباره ماهیت اساسی آنچه که آزمون عملاً اندازه میگیرد (پسند) شده است. تحقیقات بیشتر و همینطور تجربیات بیشتری در این زمینه لازم میاید قبل از آنکه یک نفر بتواند بروشنی ولی بدون هیچ ابهامی اعتبار ساختی و طبیعت ساختارهای روانشناختی را معین سازد.

۶-۶ عواملی که بر معتبر بودن آزمون اثر میگذارند

بسیاری از عوامل هستند که بر اعتبار نمرات آزمون تأثیر میگذارند. برخی از این عوامل خیلی مشخص معلومند. مثلاً، هیچکس با آزمون زبان انگلیسی معلومات ریاضی دانش‌آموزان را نمیسنجد، چرا که این آزمون آنچه را که مدنظر است نمیتواند ارزیابی کند، در نتیجه فقط به اعتبار آزمون براحتی لطمه میخورد. یا همینطور هیچکس مهارت زبانی دانش‌آموزی را با آزمونی کاملاً سخت و مشکل نمیسنجد. زیرا که توانائی واقعی وی را محک نمیتواند بزند. این به تناسب موضوعی آزمون نیز صدمه میزند. یک چنین عواملی بسادگی میتوانند تجربه شوند (مورد تجربه قرار گیرند) یا کنار گذاشته شوند بدون آنکه چندان نیازی به تبحر خاص باشد.

۶-۶-۱ دستورالعمل راهنما

دستورالعمل آزمون باید کاملاً صریح و روشن باشد تا آزمون شونده بداند که چه باید کرد. مثلاً وی باید بداند که آیا روی برگه سؤال پاسخ دهد یا روی صفحه کاغذی جداگانه میتواند سؤالاتی را که مطمئن نیستند حدسی جواب دهند، در آن صورت آیا نمره منفی خواهند گرفت. آیا اجازه سؤال کردن سر جلسه امتحان را خواهند داشت. نادیده گرفتن این نکات بظاهر بی‌اهمیت و ساده از اعتبار نمرات آزمون خواهد کاست، زیرا که نمرات اکتساب شده بدلیل همین نکات بالا و پائین خواهند شد نه بدلیل توانائی واقعی آزمون شونده.

۶-۶-۲ مشکل بودن سطح آزمون

قبلاً هم گفته شد که ساده بودن یکی از خصوصیات یک آیتم واحد است. آیتم خیلی ساده یا خیلی مشکل اعتبار آزمون را به مخاطره خواهد افکند زیرا که چنین آیتمهایی یا بالاتر یا پائین‌تر از سطح توانائی آزمون دهندگان است. و بدین ترتیب معلومات واقعی آزمون شونده را اندازه‌گیری نمی‌کنند.

۳-۶-۶ ساخت آیتما

آیتمهایی که بد تنظیم شده‌اند یا دارای ابهام هستند موجب بی‌اعتباری آزمون و مانع از بروز استعداد واقعی آزمون شونده می‌شود. اگر یک آزمون شونده آیتمی را که بد تنظیم شده از دست بدهد معلوم نخواهد بود که واقعاً جواب آنرا نمیدانسته یا بدلیل نقصی که داشته نتوانسته آنرا بفهمد.

۴-۶-۶ تنظیم آیتما و پاسخ‌های درست

معمولاً تنظیم آیتمهای یک آزمون براساس ترتیب امری مشکل است. یعنی اینکه یک آزمون نوعاً از آیتمهای ساده شروع و تا آیتمهای سخت و مشکل ادامه مییابد. علاوه بر این برای اینکه آزمون شونده رمز پاسخ دادن به سئوالات را پیدا کند آیتما براساس الگویی نامشخص و بطور نامنظم تنظیم میشوند. اگر آیتما از ساده به مشکل تنظیم نشوند، آزمون شونده در همان آغاز ممکن است دلسرد و نومید شود، و بدین ترتیب نمرات ارزش و اعتباری نداشته باشند.

با توجه به عوامل فوق‌الذکر، آزمون سازان میتوانند از پاره‌ای عوامل ناخواسته که میتوانند موجب بی‌اعتباری آزمون شوند اجتناب ورزند. هر چقدر کنترل بیشتری بر این عوامل اعمال شود، میتوان اطمینان بیشتر به نمرات بلحاظ معتبر بودن داشت.

۷-۶ پیوند بین قابل اعتبار بودن و قابل اتکاء بودن

گرچه مطمئن بودن و معتبر بودن دو مقوله‌ای بودند که بطور جداگانه مورد بحث قرار گرفتند. در اغلب موارد بطور تنگاتنگی باهم پیوند دارند. قبل از آنکه به شرح رابطه بین این دو بپردازیم دو نکته باید روشن گردد. نخست، قابل اتکاء بودن و معتبر بودن مربوط به نمرات یک آزمون است نه خود آزمون. یعنی ادعای اینکه آزمونی قابل اتکاء یا قابل اعتبار باشد عموماً یک اشتباه است. برعکس باید گفت نمرات یک آزمون مطمئن و معتبرند.

دوم اینکه، پایداری یک مفهوم ریاضی است، یعنی اینکه وقتی آزمونی درجه‌ای از قابل اعتنا بودن را نشان میدهد. بهمان میزان نمرات ثابتی را موجب میشود. اما اعتبار داشتن یک اصطلاح نسبی است. معتبر بودن بستگی به هدف آزمون دارد. بمحض آنکه آزمون از هدف منحرف گردد، اعتبار آن هم از دست میرود. مثلاً آزمونی ممکن است به یک منظوری معتبر محسوب شود اما بخاطر منظور دیگری مطلقاً اعتبار نداشته باشد یا در نظر داشتن این نکات، خواننده باید بخاطر داشته باشد که پایا بودن یک مفهوم مستقل ریاضی است. محاسبه آن، بدون توجه به شیوه‌ای که بکار گرفته میشود، کلاً متکی به یک رشته نمره می‌باشد. معتبر بودن از طرف دیگر انطباق کامل با موضوع آزمون دارد. مثلاً اعتبار داشتن برپایه معیار، مبتنی است بر بستگی بین دو آزمون که یک موضوع یا همان مطالب درسی مشابه را میسنجد. در هر صورت، باید در نظر داشت که نسبتی ریاضی مهمی بین مطمئن بودن و معتبر بودن وجود دارد.

اگر آزمونی مطمئن باشد، ممکن است معتبر هم باشد یا نباشد. مثلاً اگر یک آزمون خاص درباره درس ریاضی نتایج ثابتی به بار آورد، هرگز معیار معتبری برای توانایی زبان کسی نیست، ولو معیار معتبری برای درسی خاص ریاضی باشد. برعکس، اگر آزمونی درجه معینی از معتبر بودن را بنمایاند، تا حدی قابل اعتناء می‌باشد. بهمین خاطر هم هست که معتبر بودن مهمترین خصوصیت یک آزمون است. اگر آزمونی قابل اطمینان نباشد بستگی‌ای هم وجود نخواهد داشت. در اصطلاح فنی، مربع بستگی واریانس عادی بین دو آزمون می‌باشد. در هر آزمونی باید وجه معتبر بودن در کار باشد تا بستگی واقعیت یابد. در واقع بکمک فرمول زیر میشود این رابطه را توضیح داد.

$$r_{xy} = \sqrt{r_x} \sqrt{r_y}$$

در این فرمول، r_{xy} دلالت بر رابطه بین x و y یعنی معتبر بودن x و r_x دلالت بر قابل اطمینان بودن x و y برترتیب دارد. فرمول حاکی از این است که بستگی بین دو آزمون در بالاترین حد برابر است با حاصل ضرب ریشه مربع ضریب اطمینان دو آزمون به سخن دیگر، اگر آزمون حاکی از معتبر بودن باشد، دستکم بمیزان مربع بودن آن، آزمون دارای ضریب اطمینان است.

از بحث پیرامون رابطه بین معتبر بودن و مطمئن بودن، چنین برمیاید که میشود نتیجه گرفت که آزمون مطمئن ممکن است معتبر باشد یا نباشد، اما آزمون معتبر تا حدی میتواند قابل اطمینان باشد. بنابراین معتبر بودن خیلی مهمتر از قابل اطمینان بودن است.

۸-۶ پایایی، اعتبار و مقبول بودن

سئوالی که معمولاً از طرف دانش‌آموزان و معلمان سئوالی عادی است درباره بزرگنمایی قابل قبول و دلخواه ضریب اطمینان و اعتبار است. یک آزمون چگونه باید مطمئن و معتبر باشد؟ برای این سئوال پاسخی حاضر و آماده و باصطلاح در آستین وجود ندارد. این پاسخ تا حد زیادی به استفاده از نمرات آزمون، و اهمیت تصمیماتی که براساس این نمرات گرفته میشود دارد. به موارد زیر توجه کنید. تصور کنید یک برنامه فوق‌العاده و جدید تحصیلی که قرار است توانائی درک خواندنی دانش‌آموزان را اندازه بگیرد براساس استفاده از کتاب قرائت تنظیم شود. البته، لازم میاید تا دانش‌آموزان براساس همین مطالب امتحان گرفته شود. نتایج حاصل از یک چنین آزمونی از لحاظ تصمیم‌گیری درباره فعالیت یادگیری دانش‌آموزان نقش بااهمیتی ندارد. بنابراین، حتی اگر آزمون قدری مطمئن و معتبر در این شرایط برگزار شود، اثر جدی‌ای بر زندگی دانش‌آموزان نخواهد گذاشت. برعکس، زمانی که آزمونی برای اتخاذ تصمیمات مهمی نظیر کلاس‌بندی، برنامه‌های فوق‌العاده، دوره‌های جهشی، پذیرش برای برنامه‌های آموزشی برگزار میشود می‌بایست فوق‌العاده معتبر و مطمئن باشد. البته، تصمیم‌گیری در حوزه‌های آموزشی یک فرایند تک ساختی نیست. بندرت مربیان براساس نمره‌ای واحد در آزمونی واحد تصمیمات خطیری میگیرند. در بیشتر مواقع امر تصمیم‌گیری مستلزم مفاهیم چندگانه برای کسب اطلاعات می‌باشد. بعنوان یک قانون کلی. هر چقدر اطلاعات درباره یک مبحث بیشتر باشد، تصمیمات

منطقی‌تری پیرامون آن مبحث اتخاذ خواهد شد. بنابراین، مربیان برای آنکه تصمیمات درستی بتوانند بگیرند، بایستی به اطلاعات بسیاری دسترسی داشته باشند. به‌رحال جایی که نیاز به تصمیمات نهایی هست، مربیان ناچارند تا درصدد کسب اطلاعات مطمئن و معتبرتری بریایند. بدین‌ترتیب پاسخ بدین سؤال که در آغاز این بخش آمد بستگی به اهمیت تصمیمی دارد که بر مبنای نمرات آزمون گرفته میشود. هر چقدر این تصمیمات مهمتر باشد، احتمال اینکه نمرات مورد اطمینان بیشتری باشند زیادتر میشود. و بدین‌ترتیب، آزمون‌های مطمئن و معتبرتری مورد نیاز است. با این حال این یک قاعده عموماً پذیرفته شده است که ضریب اطمینان و اعتبار زیر $0/50$ پائین، $0/50$ تا $0/75$ متوسط و $0/75$ تا $0/90$ یا به بالا عالی است. در آزمون زبان نیز، این متغیرها بسته به شرایط مختلف که آزمون برگزار میشود میتوانند مورد استفاده قرار گیرند.

۹-۶ قابل اجرا بودن در عمل

در سرآغاز این فصل، اشاره شد که آزمون خوب بایست سه ویژگی داشته باشد، یعنی معتبر بودن، مطمئن بودن و عملی بودن. در اصطلاح اول شرح داده شد. خصوصیت سوم، یعنی عملی بودن در این بخش مورد بحث و بررسی قرار خواهد گرفت.

با وجود آنکه مطمئن بودن و معتبر بودن دو خصوصیت مهم یک آزمون می‌باشد، قابل اجرا بودن در عمل نیز مهم است. صرف‌نظر از اینکه آزمونی مطمئن و یا معتبر باشد، باید در عمل و اجرا مفید فایده باشد. بطور کلی بگوئیم که قابل اجرا بودن دلالت بر سادگی آزمون در اجرا و نمره دادن به آن دارد. علاوه بر این دو نکته مهم، عوامل دیگری مثل زمان اجرا، هزینه آزمون، سادگی ارزیابی و نمره دادن، در دسترس بودن اشکال قابل قیاس، به‌اجرای آزمون در عمل کمک میکنند.

۹-۶-۱ سادگی در اجرا

آزمون‌ها معمولاً توسط معلمان به‌اجرا درمیایند. در اغلب موارد، این معلمان آموزش خاصی پیرامون آزمون گرفتن ندیده‌اند. بنابراین، مدیریت ساده یک آزمون امر مهمی برای چنین معلمانی است. یکی از عواملی که یک آزمون را برای اجرا آسان می‌سازد روشنی و سادگی دستورالعمل آزمون می‌باشد. این دستورالعمل روشن، سهل و ساده برای خواندن، در جلسات آزمون زبان انگلیسی عموماً و زبان تخصصی خصوصاً از اهمیت فوق‌العاده‌ای برخوردار است. در چنین مواردی این دستورالعمل میتواند جهت اطمینان خاطر به زبان اصلی آزمون شونده ترجمه گردد. بویژه در مورد دانش‌آموزان مرحله مقدماتی تخصصی توجه به ترجمه دستورالعمل بزبان اصلی حتی میتواند مفید و مطلوب باشد. یک عامل دیگر که بر مشکل آزمون در اجرا میافزاید، مسئله تعداد زیادی تست‌های فرعی در آزمون می‌باشد. هر چه تعداد اینها کمتر باشد؛ اجرای آزمون در عمل ساده‌تر میشود، زیرا که شمار تست‌های فرعی بر

پیچیدگی آن اضافه میکند. و بالاخره، رعایت زمان لازم برای آزمون بدلیل احساس خستگی دانش‌آموزان از نشستن سر جلسه امتحان بمدت طولانی از اهمیت برخوردار می‌باشد.

هرگونه دستورالعمل مشکل، زمان‌بندی نامناسب، و سایر جنبه‌های آزمون در عمل که اسباب زحمت بیشتر شود، اثری منفی (معکوس) بر نتایج آزمون خواهد گذاشت و در نتیجه باعث ناکارآمدی آن در عمل خواهد شد.

۲-۹-۶ حفظ سادگی در نمره دادن

بطور سنتی یکی از مراحل پوزحمت و وقت‌گیر فرایند آزمون مرحله نمره دادن بدان می‌باشد. همانطور که در فصول پیش گفته شد، به یک آزمون به روش نظری یا عینی می‌شود نمره داد. روش نمره دادن نظری با توجه به معتبر بودن و مطمئن بودن آزمون‌ها مشکلات بسیاری فراهم می‌سازد. بعلاوه، پاره‌ای وقتها، آزمون شامل بخش‌های فرعی متعدد با روشی پیچیده برای نمره‌گذاری است. این نوع پیچیدگی کار نمره دادن را کاملاً امری کسالت بار می‌سازد.

بلحاظ آنکه نمره دادن کاری مشکل و وقت‌گیر است. گرایش به سمت عینیت، سادگی، نمره دادن با دستگاه رایج شده است. بدین معنی که آزمون سازان تا آنجا که ممکن است از نمره دادن بروش نظری احتراز می‌کنند، امروزه، بااستثنای مواردی نادر، تقریباً تمامی آزمون‌های استاندارد بطریق عینی نمره‌گذاری میشوند. گذشته از این آزمون سازان بیشتر تمایل به سیستم نمره‌گذاری خیلی ساده دارند. و بالاخره اینکه، تحولات تازه در زمینه فن‌آوری جدید موجب سهولت بیشتر در روش‌های نمره‌گذاری گشته است. با استفاده از برگه‌های پاسخ برای نمره توسط ماشین، تقریباً تمامی مشکلات نمره دادن برطرف شده است. البته، جایی که استفاده از ماشین میسر نباشد استفاده از برگه‌های جداگانه پاسخ و کارت‌های جدولی نمره‌گذاری کمک بزرگی است. بطور خلاصه، در صورتی که همه عوامل مساوی باشند آزمونی که شرایط نمره دادن و اختصار را رعایت کند بطوریکه از دقت نمرات هم هیچ نکاهد باید انتخاب و مورد استفاده قرار گیرد.

۳-۹-۶ سهولت تفسیر و کاربرد

بدیهی است که تألیف یک آزمون کاری است که هزینه دارد. هدف از آزمون تصمیم گرفتن بر مبنای جنبه خاصی از زندگی آزمون شونده می‌باشد. هیچ مهم نیست که یک آزمون تا چه حد قابل اطمینان، معتبر یا در اجرا ساده است. بلکه مهمترین نکته در مورد آن معنی‌دار بودن نمرات است که از آن اخذ میشود. مثلاً نمره X در یک آزمون به چه معنی است. آیا این نمره براساس یک سیستم قانونمند ارزیابی شده است. آیا وزن خطای معمول در اندازه‌گیری می‌باید به حساب آورده شود یا خیر؟ چه تصمیم فوری‌ای میتوان گرفت؟ این‌ها و سئوالات متعدد دیگری در راستای اجرای مؤثر و ارزیابی

درست نمرات پرسشهای اساسی‌ند. چنانچه نتایج آزمون بد تفسیر یا بد بکار برده شود، ارزش کمی دربرخواهندداشت و ممکن است در عمل بزبان فرد یا گروه تمام شود.

البته، در مورد آزمون‌های استاندارد اطلاعات مربوط به تفسیر نمرات و استفاده از آنها معمولاً از طریق کتابچه‌های آزمون در دسترس هستند. بهر صورت باید دقت شود تا دید که تبدیل نمرات خام به نمرات معنی‌دار چگونه است و جدول معیارها تا چه اندازه بروشنی ارائه شده است و به چه میزان توصیه‌هایی دایر بر بکارگیری نتایج بطور کامل و جامع به عمل آمده است.

در یک جمع‌بندی در کنار معتبر بودن و مطمئن بودن یک آزمون، ملاحظات عملی نقشی مهم در استفاده از آزمون بازی میکنند. بطور کلی، عملی بودن با تمهیداتی ارتباط دارد که نیازمند اجرای یک آزمون است. این تمهیدات عبارت است از روش نمره دادن و برگزاری آزمون مقدمات لازم، و تفسیر نمرات اخذ شده، به لحاظ آگاهی محدود اکثر معلمان مدارس در زمینه تکنیک‌های لازم در فرایند آزمون، باید تلاش شود تا آزمون‌هایی تنظیم، انتخاب یا از آنها استفاده شود که براحتی قابل اجرا، ارزیابی و نمره دادن باشند.

بهر ترتیب، باید گفت که گرچه قابلیت اجرا در عمل یک عامل مهم در انتخاب و استفاده از یک آزمون است. مربیان نباید از اهمیت و اعتبار تصمیماتی که بر مبنای نمرات آزمون میگیرند غافل بمانند. در صورت اتخاذ تصمیماتی پراهمیت، قابلیت در اجرا نیابستی چندان عامل تامین کننده‌ای تلقی شود. مثلاً استفاده از موضوع انشاء از نقطه نظر نمره دادن چندان قابلیت عملی ندارد. بخصوص بهنگام برگزاری جلسات امتحانی بزرگ و همینطور اجرای مصاحبه‌های شفاهی. بهرحال، زمانی که انتخاب معلمان زبان در دستور کار قرار دارد، مشکلات مربوط به اجرا و نمره دادن در مصاحبه شفاهی نیابستی مسئولان را بترساند، زیرا که زبان گفتاری برای معلمان زبان برای اینکه خودشان را معرفی کنند بشناسانند از اهمیت فوق‌العاده‌ای برخوردار است. یا زمانی که گزینش نویسندگان حرفه‌ای هدف آزمون است. برگزاری یک آزمون متن از واجبات است. زیرا که در انتخاب چنین افرادی، نوشتن نقش اساسی دارد. البته مرحله نخست در گزینش در آزمون‌هایی با مقیاس بزرگ با همان قلم و کاغذ شروع میشود. این روش از شمار متقاضیان میکاهد. پس در مراحل نهائی و با تعداد اندک آزمون دهنده، شیوه‌های دیگری صرف‌نظر از قابلیت‌های عملی آنها میتواند بکار آید.

فعالیت‌ها

۱- آزمونی مربوط به درک مطلب خواندنی واریانس برابر ۲۵ و ضریب اطمینانی برابر ۰/۸۱ دارد، خطای معمولی این اندازه‌گیری را محاسبه کنید.

۲- حداکثر بستگی بین دو آزمون با ضریب اطمینان ۰/۶۴ و ۰/۸۱ چیست؟

۳- به چه دلیل شیوه اشکال هم‌تراز (همسو) در محاسبه پایایی (ضریب اطمینان) بر روش آزمون پشت آزمون برتری دارد؟

- ۴- آزمون واژگان ۴۰ آیتمی دارای میانگینی برابر IO و SD برابر ۱۰ دارد، ضریب اطمینان این آزمون چقدر است؟
- ۵- ضریب اطمینان آزمونی از طریق روش دو نیمی ۰/۷۰ شده است، پایایی تمامی آزمون چند است؟
- ۶- چطور و چرا یکدستی موضوع درسی و آیتمها بر ضریب اطمینان و اعتبار یک آزمون تأثیرگذار است؟
- ۷- یک آزمون ۲۰ آیتمی دارای یک ضریب اطمینان ۰/۴۰ می‌باشد اگر شماره آیتمها به ۳۰ افزایش یابد، ضریب اطمینان آزمون جدید چقدر خواهد بود و همینطور اگر ۴۰ آیتمی شود؟
- ۸- معتبر بودن آزمون همزمان چه تفاوتی با معتبر بودن آزمون تخمینی دارد؟
- ۹- کدام بهتر است؟ اطمینان داشتن یا اعتبار داشتن؟ چرا؟
- ۱۰- رابطه بین خصوصیات آیتمی و آزمونی در چیست؟

آزمون زبان و نظریه‌های درباره آن

مقدمه

در فصول پیش، نقش آزمون‌ها، ویژگی‌های یک آزمون خوب، فنون مختلف طراحی سؤال برخی مفاهیم اساسی آماری مورد بحث و بررسی قرار گرفت. این فصل با طرح موضوع آزمون زبان یکبار دیگر به رئوس مطالب بحث‌های پیش خواهد پرداخت. بدین منظور، شرح و ارزیابی مختصری از نظریه‌های گونه‌گون پیرامون آزمون‌سازی و بررسی نمرات آن بدست خواهیم داد.

طراحی آزمون زبان و تحول زبان‌شناسی مدرن، معلمان و طراحان آزمون را به اهمیت ضرورت یک تحلیل جامع درباره زبان موردنظر واقف ساخته است. معلمان از طرفی پیوسته به تجدیدنظر در راهکارهای آموزشی خود در پرتو پیشرفتهای تازه در زبان‌شناسی جدید پرداخته‌اند. و از طرفی دیگر طراحان در تلاشند تا همگام با پیشرفت در امر آموزش فنون آزمون‌های زبانی را بگونه‌ای مطمئن‌تر و معتبرتر به پیش برانند. به یقین تمامی نوآوری‌ها در حوزه زبان تأثیری یکسان و مشابه بر آموزش و آزمون زبان ندارند. بلکه هر کدام از آنها بنوعی بر اهمیت نسبی هر یک از مهارت‌ها یا مؤلفه‌های زبان تأکید می‌گذارند.

امر سنجش مهارت‌های زبانی مستقیماً برمیگردد به اینکه واقعاً هدف از دانستن زبان چیست. سئوالی که مطرح می‌باشد اینست که آیا زبان میتواند به مهارت‌ها و مؤلفه‌های آن تجزیه گردد، آیا به تمامی مهارت‌ها بطور جداگانه و بمیزانی متفاوت یا باهم و به یکسان میتوان رسید. بعنوان نتیجه، نقطه نظرات مختلف پیرامون زبان و اتکاء احتمالی آنها به گرایشات متنوع در آموزش زبان، چهار گرایش عمده همانند، در امر طراحی آزمون و در رقابت باهمند. تحولاتی که در نظریه زبان‌شناسی پدید آمده، از

نظریه سنتی گرفته تا ساختاری، تحولگرا تا برسیم به نظریات عمل گرایانه در علم کلام (زبان) - و نقطه نظرانی منبأ آموزش و یادگیری، طراحان آزمون را واداشته است تا بیشتر بفکر کیفیت خود آزمون باشند تا تنها بدنال حداکثر نشانه‌های روانسنجی. جدول ۱-۷ نمودار ساده‌ای از طبقه‌بندی این مشخصه‌ها با توجه به آغاز پیدایش آنها می‌باشد.

جدول ۱-۷- طبقه‌بندی مشخصه‌ها با توجه به آغاز پیدایش آنها

روانشناسی	زبان‌شناسی	تدریس	آزمون‌سازی
	سنتی	ترجمه - گرامر	ترجمه - گرامر
رفتارگرایی	ساختارگرایی	شنیداری - گفتاری	نکات مجزا
شناخت‌گرایی	زایشی	رمز - شناختی	کامل (یکپارچه)
روان‌زبان‌شناختی	عملکردی	نقشی - مفهومی	نقش عملکردی

آزمون زبان تا حال از خصوصیت آموزش زبان که به نوبه خود تحت تأثیر اصول زبان‌شناسی و روانشناسی بوده تبعیت کرده است. مثلاً آزمون زبان پیش از آنکه با مطالعه زبان بعنوان یک علم کلامی شناخته شود، بخاطر فقدان یک نظریه تثبیت شده چنین شده است. مسئله‌ای بنام عینی یا پیاپی وجود نداشت. آزمون‌ها عبارت از ترجمه متن، تمرینات کتبی کامل کردنی و انشاء بودند. این شیوه غیرعلمی ناشی از اتخاذ روش گرامر - ترجمه در امر آموزش بود، که صرفاً بر شکل صحیح در خواندن و نوشتن اتکاء میکرد. روش گرامر ترجمه در دهه سال ۱۹۵۰ حاصل گرایش دیرینه در تحلیل زبان‌شناسی‌ای بود که میل به سنجش سطحی و روینائی داشت. با این وجود تا حال بلحاظ تأثیر مقوله رفتارگرایی، در روانشناسی و ساختارگرایی در زبان‌شناسی، معلمان زبان از طریق ایجاد عادت در سود بردن از الگوی زبانی به شیوه شنیداری در آموزش روی آورده‌اند. دانستن زبان بمنزله دانستن مؤلفه‌ها و اجزای آن، قادر بودن به استفاده از آن در مهارت‌هایی نظیر شنیدن، صحبت کردن، خواندن و نوشتن می‌باشد. این مؤلفه‌ها در قالب مهارت‌ها چهارچوبی در اختیار می‌گذارد تا آزمون سازان اساس آزمون خود را بر آن بگذارند. فرض بر این است که آزمون‌ها طوری تحول یابند که بتوانند معلومات درون هر یک از این خانه‌ها را از طریق کارت الگوی جدولی بسنجند.

با توجه به زبان‌شناسی که حاصلش فرایند بیان است و روانشناسی که منجر به شناخت میشود، معلمان زبان روش یادگیری رمز - شناختی را ابداع کرده‌اند. در نتیجه استفاده از الگوی جدولی بر این پایه مورد انتقاد قرار گرفته که با ذره ذره‌ها نمیشود توانائی فرد را در زبان اندازه گرفت. و بجای آن تکنیک‌های آزمون کامل توصیه شده است.

علاوه بر این، از آنجا که نظریه زبان‌شناسی بر اصل عمل‌گرایی و روانشناختی تئوریهای یادگیری، بویژه نظریه یادگیری / تسلط بر یک زبان دوم / خارجی استوار است شیوه‌های آموزش بر روش متحول

توانمندی در گفتگو و مکالمه متمرکز شده است. در عوض آزمون‌ها هم بمنظور سنجش همین مهارت‌های گفتاری تحول یافته‌اند.

آنچه که در زیر می‌آید، تشریح و ارزیابی کوتاهی است از سه نظریه مورد قبول همگان. روش ترجمه - گرامری در اینجا بدلیل آنکه در آموزش زبان ما را به هدف نمیرساند مورد بحث قرار نگرفته است.

۲-۷ روش نکات تجربیدی (مجزا شکل)

آزمون‌ها بروش سنجش نکات تجربیدی در ارزیابی توانایی‌های زبانی از طریق موضوعات مختلفی که نمونه مؤلفه هر یک از مهارت‌ها، وجوه گونه‌گون و اجزای آنهاست برمی‌ایند. دلیل منطقی برای استفاده از آزمون‌های تجربیدی (DP) ناشی از تعامل بین زبانشناسان ساختارگرا و روانشناسان رفتارگرا است. روانشناسان قواعد سنجش آموزشی را ارائه داده‌اند و مدعی‌ند که یادگیری از طریق ایجاد عادات مربوطه ممکن است، و زبانشناسان لیست بالا بلندی از آیتمهای کوچکی که از تجزیه زبان به واحدهای کوچکتری حاصل آمده است در دست دارند. اینها به معلمان زبان در تحول بخشی به شیوه‌های شنیداری در آموزش زبان و همینطور به طراحان آزمون در تحول بخشی به آزمون‌های چند گزینه‌ای کوتاه که پایائی و معتبر بودن آنها براحتی مورد تأیید قرار گیرند کمک می‌کنند گفته میشود که فقط یک مؤلفه بایستی در یک زمان مورد آزمون قرار گیرد، که تنها وجهی از یک مهارت بایستی در یک زمان سنجیده شود. آیتمهایی که بمنظور سنجش واج‌های صوتی زبان آزمون‌دهنده تنظیم میشوند، یا الگوهای زیر و بم‌دار صدا، واژگان، استراکچر و نظایر آن بخشی از همان روش مجزا شکل می‌باشند. جدول ۲-۷ نموداری از یک جدول مهارت‌های دو وجهی در یک آزمون تواناییهای زبانی است. آزمون‌های تنظیم شده براساس این نظریه در فصول بعد ارائه خواهد شد.

شیوه DP^۱ در تنظیم و تهیه شمار زیادی از آزمون‌های استاندارد و دقیق نظیر تافل^۲ که در دنیای امروز معیار قابل اعتمادی برای تشخیص توانایی زبانی است بکار گرفته میشود. در واقع شیوه DP میرود تا به شیوه‌ای عمده در آزمون سازی در سال‌های آینده درآید. آزمون‌های طراحی شده بدین روش هم مورد تحسین و هم مورد نقد قرار گرفته‌اند. انتقاد از آن متوجه ناکارائی شیوه‌های زبانشناختی و روانشناختی‌ای میشود که اساس DP بر آن نهاده شده است. مثلاً چنین روشی طراح آزمون را قادر می‌سازد تا توانائی فرد را در بکارگیری وجهی از زبان در شکل یکی از چهار مهارت اصلی بطریق ایدال بسنجد. در این معنی آزمون‌های DP شکل ایدال آزمون‌هایی با هدف تشخیصی می‌باشند. با این وجود، همانطور که در فصول بعد نشان خواهیم داد، تنظیم آیتمهای DP محض فوق‌العاده مشکل است.

^۱ - Discrete point

^۲ - TOEFL

بسیاری از آزمون‌ها که بر این روش متکی هستند شامل آیت‌هایی است که بر پایه‌ای بیشتر از یک وجه یا مهارت عمل میکنند.

مزیت متمایز آزمون‌های DP این است که نتایجی را به‌بار می‌آورد که خیلی راحت میشود گفت کمی است، گرچه به سختی میشود ثابت کرد که چه ارزشی دارد. آزمون‌های DP با این اعتقاد طراحی میشوند که دانستن درباره زبان یعنی دانستن خود زبان بهر صورت آگاهی بر عناصر یک زبان ارتباط کمی یا اصلاً ربطی به زبان در زندگی واقعی ندارد. منتقدان بر این باورند که الزام به انتخاب کردن و نکردن نکاتی برای امتحان، کار پر زحمتی است. گرچه تحلیل افتراقی بعنوان وسیله‌ای برای گزینش نکات آزمونی بکار میرود، بخاطر عدم دوراندیشی مورد انتقاد قرار گرفته است. علاوه بر این تهیه آزمونی که نمونه الگوی‌های زبانی یک دنیا باشد عملاً غیرممکن است، گذشته از این که تازه قادر به سنجش توانایی فرد در استفاده از زمان در زندگی واقعی هم نیست.

بدلیل یک چنین نارسائی‌های نظری و عملی‌ای است که تمامی توجه به روش‌های جایگزین جلب شده است آزمون‌های جامع بعنوان معیارهای برتر سنجش پیشنهاد شده است؛ این آزمون‌ها همراه و هماهنگ آزمون‌های DP عمل میکنند.

۳-۷ روش جامع

آزمون‌های جامع برآنند تا جنبه‌های عملی کارهایی که فرد باید بطور معمول به‌نگام استفاده از زبان انجام دهد بسنجند. دلیل منطقی تاکید بر آزمون‌های جامع اینست که چون منظور از زبان گفتگو است، پس آزمون زبان هم بایست براساس کارکرد گفتاری زبان در شرایطی مشابه، وضعیت زندگی واقعی تنظیم گردد. برخلاف آزمون‌های DP، آزمون‌های جامع به اشکالی در زبان وابسته‌ند که همه گزینه همدیگرند. یعنی اینکه، بعوض این تلقی که زبان مجموعه‌ای از عادات مختلف است، روش جامع زبان را بعنوان یک سیستم دینامیک که تمامی مهارت‌های فرعی نیز در یک زمان مورد ارزیابی قرار میگیرد، میشناسد. بنابراین، آزمون‌های ساخته و پرداخته، بر این اساس، توانایی آزمون‌دهنده را از طریق پویایی فرایند اطلاعات ارائه شده محک میزند. یک چنین دلیل منطقی‌ای حاصل تمرکز زبان‌شناسان تحول‌گرا بر توانایی زبانی و روش منطقی روانشناسان شناخت‌گرا در تلاش برای کشف اصول روانشناختی است تنظیم و اجرای جاهای خالی، دیکته، نوشتن متن، مقاله، مصاحبه شفاهی نمونه‌هایی از آزمون‌های جامع می‌باشند.

البته، هر کارشناسی با این نظم موافق نیست که آزمون‌های جامع با آزمون‌های نکات تجریدی متفاوت عمل میکند. برخی اعتقاد دارند که آزمون‌های جامع و نکات تجریدی هر دو نشانه شکل یکسانی هستند؛ تنها به‌لحاظ اندازه متفاوتند. یعنی اینکه، این کارشناسان اعتقاد دارند که آزمون‌ها جامع جمیع جهات و نکات تجریدی دو نقطه در انتها الیه یک مسیرند، در یک سو آیت‌های مجزاترین

شکل نکات: و در سوی دیگر، جامعترین آزمون‌ها با انواع بسیاری آیتم (نظیر آیتم‌های درک مطلب خواندنی) که در این فاصله قرار دارند (ما بین).

منتقدان در آزمون جامع، نارسائی بسیاری مشاهده میکنند. در فضل پانزده پاره‌ای معایب ذاتی این نوع آزمون را برمیشماریم. خلاصه اینکه آزمون‌های وقت‌گیر و نمونه‌ای نظیر انشاء نویسی بلحاظ برگزاری و نمره‌گذاری چیزی و رای ملاحظات عملی آزمون‌های از حد بزرگ است. بعلاوه، ضریب اطمینان و پایائی جاهای خالی در آزمون دیکته محل سؤال می‌باشد.

و بدین ترتیب معتبر بودن آنها در معرض ارزیابی، غلط است، بهمین خاطر تست‌های جاهای خالی و دیکته مطابق با معیار پیش‌بینی یا ارزیابی یک مهارت کامل و جامع زبانی نیست.

با توجه بیشتری که زانشناسان به جانب روش‌های معناگرایانه و نه صرفاً صرف و نحوی معطوف داشته‌اند، پاره‌ای از معلمان دلیل می‌آورند که آزمون زبان باید کارائی مورد انتظار گرامری آزمون دهنده را بسنجد. یک مهارت گرامری مورد انتظار بلحاظ ذهنی (روانی) آن چیزی است که دانش‌آموخته بنام گرامر واقعی بهنگام یادگیری زبان درونی خود می‌سازد. ابزار اندازه‌گیری‌ای که برای سنجش این کارائی بکار می‌آید اصول عملکرد شناسی است. به سؤالات پاسخ دادن، دیکته نوشتن، قرائت، تست Noise، ترجمه، انشاء نویسی، مصاحبه شفاهی (حضوری)، بازگوئی متن و تشخیص آن، نمونه‌هائی از آزمون سازی عملگرایانه به حساب می‌آید.

اصولاً یک اصل گرامری متکی به انتظار شباهت به گمانه‌زنی‌های مطمئنی درباره آنچه که گوینده گفته است دارد و همین‌طور به شخص آزمون شونده درباره آنچه که بعداً قرار است بگوید. از نقطه نظر شنونده، گمانه‌ها، نظریاتی هستند دال بر آنچه که خود وی تا حال فهمیده است و اینکه گوینده بعد از این احتمالاً چه خواهدفهمید. مشابه همین قضیه از نقطه نظر یک خواننده و نویسنده بودن اتفاق می‌افتد. بسخن دیگر انتظاری که از وضعیت گرامری می‌رود بی‌شباهت به شکل گرامری نیست که گوینده در صحبت‌های خودش بدان احتیاج مبرم دارد. بر همین اساس آزمون‌های مبتنی بر اصل عملگرائی برآنند تا توانائی تشخیص رابطه بین مؤلفه‌های زبانی و شرایط غیر زبانی که مؤلفه‌ها در آن بگونه‌ای مناسب بکار می‌روند بسنجند. یک آزمون وقتی از اصل عملگرایانه تبعیت میکند که دو معیار را برآورده سازد. اول شکل دادن به سیستم متحول گرامری فرد و به چالش گرفتن آن که پردازش انتظام عناصر زبانی را به اقتضای شرایط در یک آن، ضروری می‌سازد و نیز آزمون باید دارای همان نظام زبانی باشد که به شرایط فرازبانی به روشی معنی‌دار ربط یابد.

با این تفصیل، آزمون‌های عمل‌گرا همیشه یک تمامیت کمال یافته‌اند. اما علتی آن درست نیست. تست‌های اصطلاحاً جامع ممکن است همین‌گونه باشند. آنچه که بعضی از آزمون‌های جامع را تابع این اصل می‌سازد. همان حدی است که آزمون شونده میتواند در مکالمه دخالت کند. در شرایط واقعی در هیچ حالتی بهنگام صحبت کردن از آزمون شونده خواسته نمی‌شود تا به یکی دو نکته بی‌اهمیت مربوط به اصوات که چندان فرقی به حال کسی نمیکند حواسش جمع باشد. در صورتی که آزمون دهنده

تشویق گردد تا با آنچه که در یک تست جا خالی آمده بعنوان یک متن محاوره مرتبط و پیوسته برخورد کند و نه یک سری جملات بی ربط، یک تست عمل‌گرا به حساب خواهد آمد. البته تمامی آزمون‌های کامل به هر دو معیارهای فوق بعنوان تست‌های عملگرا (واقعگرا) عمل نمی‌کنند. اگر تستی تنها تسلط به یکی از چهار مهارت را مدنظر داشته باشد، تستی کامل است. حفظ کردن مطالب بدون توجه به معنای آنها نشانه آزمون واقع‌گرایانه نیست و همینطور سرهم کردن یک سری مؤلفه‌های زبانی در عمل بدون وقوف بر معنای آنها نشانه یک کار واقع‌گرایانه نمی‌باشد. برای آنکه تستی واقع‌گرایانه تلقی شود به چیزی بیشتر از این نیاز هست.

نظیر روش‌های گذشته، تست‌های واقعگرا محدودیت خاص خودشان را دارند، و اکثر آنها مهارت گفتاری زبان را در عمل محک نمی‌زنند. اعتقاد بر این است که توانمندیهای ارتباطی و مخابراتی آنچنان جهانی گشته که دیگر تست‌های نوعاً واقعگرا بقدر کفایت نمی‌توانند آنها را اندازه بگیرند. آزمون‌های عملگرا صرفاً از آزمون شونده می‌خواهند تا اغلب همزمان بر وجوه مختلف سیستم زبان و تعاملات آن در عمل تسلط یابند، با این حال مهارت گفتاری در زبان را نمی‌سنجند. مثلاً تست جاهای خالی (Cloze) تست کارکردی زبان است ولی از آنجا که شرح غیرمستقیم معلومات زبان آزمون شونده در هنگام صحبت است، مکالمات افراد را در زندگی واقعی نشان نمی‌دهد. این آزمون بلحاظ دیکته نیز مورد انتقاد قرار دارد. چرا که دلیلی متقاعد کننده دایر بر مهارت فرد مبنی بر استفاده از زبان در صحبت کردن بدست نمی‌دهد. یعنی اینکه Cloze و تست دیکته، گویای داشتن یا نداشتن مهارت در صحبت در چهارچوب استفاده از مهارت‌های زبانی در عمل نیست.

۴-۷ روش عملی

تا همین اواخر، کارشناسان زبان به اشکال زبانشناسی و تشریح ساختاری زبان توجه داشتند، اکنون این توجه به تشریح معنایی معطوف گشته است و به زبان بعنوان بخشی از کنش ارتباطی جمعی در میان آحاد اجتماعی نگریسته میشود. اکنون زبانشناسی با تحلیل گفتمان انسانها و تعاملات متقابل ارتباطی سروکار دارد. تحولی مشابه در رویکرد به روانشناسی پدید آمده است. روانشناسان بجای این تلقی از یادگیری زبان که ضمن فرایند آن نوآموز بتدریج قواعد گرامری را در حافظه خود انبار میکند و بعد بگونه‌ای می‌آموزد که چگونه از این قواعد در مکالمه زبان استفاده کند، اکنون به این نقطه نظر رسیده‌اند که زبان مهارتی است که افراد بخاطر صحبت و برای برقراری ارتباط یاد میگیرند و بعد از درون آن ساخت گرامری شکل میابد، بصیرت و آگاهی حاصل از تحلیل گفتمان دو طرفه، برای خود شخصیت قائل شدن و گفتمان (درونی) با خود انگیزه اصلی ظهور نظریه روان - زبانشناختی فراگیری / یادگیری زبان دوم / خارجی نیز هست.

تحولات پدید آمده در این دو حیطه موجبات درک تازه‌ای از اهداف تدریس زبان شده است. اکنون مکالمه زبان بعنوان حاصل و بخش مهم و الزامی تدریس زبان توجهات زیادی را بخود جلب نموده

است. از آنجا که توانش زبانشناختی بخشی از توانش گفتمانی است، روشن شده است که توجه به حوزه روانشناختی بخش‌های وسیعی از حوزه زبانشناسی را به ناگزیر دست نخورده باقی نهاده است. توانش ارتباطی - شامل توانش زبانشناختی، جامعه زبان شناختی و توان برنامه‌ریزی عبارت است از آگاهی پیرامون شکل زبان و چگونگی استفاده بجای آن در شرایط واقعی است. بر همین مبنا - تدریس زبان به جانب مکالمه کردن هدایت شده است. زیرا که مکالمه زبان خود به خود توانش زبانشناختی را هم تحول میبخشد.

بسط و توسعه منطقی شیوه مکالمه در تدریس زبان نشان میدهد که سنجش توانائی مکالمه باید هدف نخست آزمون زبان باشد. در نتیجه طراحان آزمون به دنبال روش‌هایی میگردند که بتواند این توانش گفتاری را بیازماید، یعنی اینکه مهارت در مکالمه بمفهوم نظری تمامی آن چیزی شناخته میشود که معلومات زبان خوانده میشود. و از آزمون‌دهنده انتظار میرود برای اینکه بتواند صحبت کنند یک سری تکالیف را انجام دهد. بخصوص اینکه کار آزمون‌های شفاهی - گفتاری ارزیابی مهارت آزمون دهنده در مکالمه زبان در شرایط معینی است. طی آزمون متنی با تمامی دقایق آن در اختیار آزمون شونده گذاشته میشود تا با آن به تعامل بنشیند و با توجه به تمامیت آن و نه تک تک عبارات یا جملات پاسخی شفاهی یا کتبی تهیه کند.

آزمون شفاهی - گفتاری مهارت‌های زبانی را - زمینه‌ها و شرایطی را که تحت آن این مهارت‌ها موضوعیت می‌یابد - براساس بررسی دقیق نیازهای آزمون شونده تعیین میکند. بدیهی است که آزمون با رشد و تحول برخی مهارت‌ها سروکار دارد. نه کل توانش شفاهی - گفتاری که کاری غیرقابل تصور است. یک چنین گرایشی متضمن ایجاد ملاک‌های معتبر است. چرا که تألیف آزمون بر ملاکی از پیش تعیین شده استوار است. در واقع، زبان بمنظورهای مختلفی آموخته میشود. و معیار آزمون نیز به همین منظورها متفاوت است. پس از این‌رو، مهمترین ضرورت برای آزمون زبان تشخیص ضرورت‌های گفتاری (ارتباطی) زبان آموزان می‌باشد. از آنجا که دیگر به زبان به چشم یک تمامیت یکپارچه با الگوهای تثبیت شده گرامری و اندوخته‌ای از واژگان نگریسته نمیشود، بلکه بیشتر سیستمی است با ترکیب متفاوتی از مهارت و تسلط بر زبان با هدف ارائه اشکال مختلف ارتباطی. آزمون زبان باید با نیازهای خاص افراد پیوند داشته و بدین ترتیب ویژگیهای هرگونه استفاده از زبان را به حساب آورد. بطور اختصار باید گفت خصوصیات عمده استفاده از زبان به شرح زیرند.

- ۱- کارکرد زبان مبتنی بر تعامل متقابل است که بمنظور صورت میگیرد
- ۲- تحول در تعاملات غیرقابل پیش‌بینی است؛ و نیازمند زمان واقعی است تا جا بیافتد.
- ۳- کارکرد زبان در شرایط واقعی صورت میگیرد که به نوبه خود اشکال زبان را معین میکند.
- ۴- شرکت‌کنندگان در یک گفتمان موفقیت یا شکست یک تعامل را با توجه به نتایج رفتاری آن داوری میکنند.
- ۵- توانمندی گفتاری در عمل به استعداد پردازش زبانی روان و دقیق مربوط میشود.

شاید ما از آزمون‌هائی که بطور سیستمیک این خصوصیت را مورد توجه قرار دهد خیلی فاصله داشته باشیم. اما بهر حال اینها اهداف اولیه آزمون‌های شفاهی - گفتاری هستند. شرح آزمون‌های شفاهی - گفتاری و تحول آزمون نمونه ESL برای دانش‌جویان در فصول بعد بتفصیل خواهد آمد.

بطور اختصار، کارشناسان در حال تجربه یک همگرایی جدید و واقعی‌ند. نیاز به بحث نیست که آزمون زبان بایست کارکرد زبان را ارزیابی کند و اینکه روش شفاهی - گفتاری چهارچوب نظری ایدالی برای این کار عرضه میکند. با این وجود یک دستورالعمل سیستماتیک تفصیلی برای استفاده امروزی در تألیف آزمون که مستلزم آنست در عمل امری کاملاً غیرمحتمل بنظر میرسد. بویژه تهیه و تنظیم آزمون‌های جداگانه برای هر الگوی گفتاری کاری شدنی نیست. راه حل پیشنهادی برای چنین کاری عبارت است از سیستم آزمون‌های دو وجهی: تست usage برای سنجش توانائی آزمون شونده در نوشتن جملات درست و تست use برای تشخیص میزان تسلط وی در ادای جملات بمنظور انجام یک مکالمه به دیگر سخن دو گونه آزمون برای ارزیابی دقیق مهارت کلامی موردنیاز می‌باشد. نوع اول، برای سنجش مهارت زبانشناختی کلیه داوطلبان نوع دوم که باید شامل یک سری تست‌های فرعی برای سنجش کار کلی هر آزمون دهنده‌ای باشد.

لازم به یادآوری است که در اکثر موارد وجه تفاوت بین نظریه‌های مختلف بخاطر مسئله نوع نیست، بلکه بخاطر درجات آن است. یعنی، آزمونی از یک آزمون دیگر کاملتر است؛ یا آن یکی از این مجزاتر است. بنابراین، نمیشود گفت که آزمونی کاملاً مجزا یا کاملاً جامع است. آنچه که در هر حال گفتنی است این است که تمامی تست‌ها را میشود در یک سیر منطقی قرار دارد. سؤالات تجریدی در یک طرف و سؤالات کارکردی‌تر در طرف دیگر، مثلاً، آزمون‌های واژگان تجریدی‌تر از گرامر هستند. بهمین سان، آزمون مهارت‌های زبانی انسجام یافته‌ترند تا مثلاً آزمون‌های مؤلفه‌های زبان. در واقع مهارت‌های زبانی آمیزه‌ای از سؤالات تجریدی و غیرتجریدی هستند. درجه تجریدی بودن طبق نمودار زیر ترسیم شده است.

برای آنکه توضیحات مفصلتری در اختیار دانش‌جویان قرار دهیم، در فصول بعد به شرح آزمون‌های نکات گرامری، مهارت‌های زبانی، روش‌های Holistic (ترکیبی) بطور جداگانه خواهیم پرداخت.

۵-۷ تفسیر نمرات آزمون

یک مسئله نظری دیگر در حال حاضر در زمینه آزمون زبان به چگونگی تفسیر نمرات آزمون مربوط میشود. پس از آنکه آزمونی برگزار شد و نمرات محاسبه گردید، مسئله اساسی این است که نمرات برایمان چه معنائی خواهند داشت. البته در اینجا مسئله به هدف آزمون و در نتیجه روش‌های تألیف آن مربوط میشود.

همانطور که در فصل چهار خاطر نشان شد، نمرات خام نمیتوانند بطور معنی‌داری تفسیر گردند. اعلام اینکه داوطلبی نمره ۶۰ در آزمون تخصصی گرفته است اطلاعات چندانی یا اصلاً اطلاعاتی

درباره توانایی زبان وی نمیدهد. و نه نمرات درصدی راه حل متقاعد کننده‌ای در جهت تفسیر نمرات امتحانی در اختیار میگذارد. برای رسیدن به نتایج این تفسیر دو روش ابداع شده است: تفسیر متعارف و تفسیر ملاکی. چنانچه نمره داوطلبی را با نمره داوطلبان دیگری مقایسه کنیم این تفسیر متعارف خواهد بود. در صورتی که اگر ما نمراتی را با معیاری معین قیاس نمائیم بدون آنکه بخواهیم بدانیم دیگران چه کرده‌اند، این تفسیر ملاکی خواهد بود.

نمرات امتحانی میتوانند بهر دو شیوه مورد ارزیابی قرار گیرند، برحسب قصدی که داریم، مطمئناً یکی از این راهها از دیگری مناسبتر خواهد بود. یعنی اینکه، تصمیم مبنی بر اینکه کدام روش را برگزینیم تصمیمی واقع‌گرایانه خواهد بود. هر کدام از این دو روش که انتخاب گردد، معیار اندازه‌گیری و تحلیلی خاص خودش را خواهد داشت. در زیر موارد استفاده هر یک از دو روش متعارف و ملاکی به اختصار مورد بحث و بررسی قرار خواهد گرفت.

۱-۵-۷ تفسیر متعارف

تفسیر متعارف با مقایسه نمرات داوطلبان با نمرات منتخبین در سطح محلی منطقه‌ای یا ملی به نمرات خام معنی بخشیده میشود. نتایج امتحانی در انطباق با نمرات متعارفی که نشانه عملکرد امتحانی یک جمع نمونه است که بعنوان نمونه معمولی standard انتخاب شده مورد ارزیابی قرار میگیرد. نمرات متعارف براساس آمار همان معدل نمرات امتحانی یک گروه برگزیده می‌باشد.

ملاک‌های برآمده از یک چنین روشی دارای ویژگیهای چندی است نخست اینکه حداکثر تمیز را بین آزمون شوندگان ایجاد کند و پراکنش نمرات را در سرتاسر یک منحنی فرضی معمولی موجب شود. برای اینکار و بکمک تمهیدات مختلف میشود کاری کرد که آزمون نمودار دوره یا اهداف آن باشد. در واقع، هیچ مهم نیست که آزمون شونده کدام یک از سئوالات را از دست میدهد. اما وقتی شمار زیادی سئوالات را نمیتوانند پاسخ گویند آن سئوال از تست حذف میشود. بر مبنای یک چنین عواملی مثل سادگی یا وجه افتراق سئوالات است که پیراش آزمون برقرار میشود تا متضمن این امر باشد که افراد متبحر قوی‌تر در آزمون نمرات بالاتری از ضعیف‌ترها کسب کنند.

چندان جای حرف ندارد که روش تفسیر متعارف دارای برخی مزیت‌هاست. مثلاً در تست استعداد یعنی آزمونی که منظور از آن پیش‌بینی افتراقی استعدادهاست؛ افرادی با نمرات عالتر شایستگی ادامه برنامه‌ای را بیشتر از کم استعداده‌ها دارند. در ضمن، این امر در آزمون‌های پیشرفت‌ی نظیر امتحان بخاطر توزیع نمرات پایان دوره مفید است. دانش‌آموزان زرنکتر شایستگی دریافت نمرات بالاتر دوره را پیدا میکنند. مصداق بیشتر استفاده از روش تفسیر متعارف در تست‌های گروه‌بندی است که هدف از آن توزیع آزمون دهنده از لحاظ توانایی در سطوح مختلف گروهی کلاسی است.

نوعاً آزمون‌هایی که در کلاس بکار میروند از این گونه‌اند. ملاک‌های معمولی عموماً براساس هنجارهایی که بر طبق انجام آزمون افراد در امتحانات قبلی تنظیم و تحول یافته‌اند استوارند تا کاربران

بتوانند نمرات افراد را باهم مقایسه کنند. بمنظور تنظیم قواعدی برای آزمون‌ی جدید، آزمون در اختیار یک گروه نمونه معمولی منتخب قرار میگیرد. داده‌های مربوط به نمرات اخذ شده آنگاه در جدولی که مبنای تصمیم‌گیری / پیش‌بینی درباره کار آینده افراد است درج می‌گردد. برای سهولت دو تفسیر نمرات آزمون، جدولی دارای ستون‌های درصدی است. مثلاً نمره آزمون‌ی با عدد ۴۱ در جایی بین ۸۸ و ۹۰ در جدول قرار دارد.

رتبه درصدی و نمرات معمولی درباره وضعیت نسبی یک آزمون شونده در ارتباط با یک گروه متعارف مشخص اطلاعاتی را در اختیار می‌گذارند. از این نظر هر دو مفیدند. رتبه‌های درصدی توسط آزمون دهنده براحتی قابل فهمند. اما نمرات معمولی بدلیل آنکه بیانگر نمرات یکسان هم هستند مفیدند. با همه این حرفها هر دو روش با توجه به ملاحظات زیر می‌باید بکار گرفته شوند. هر هنجاری به نمونه ویژه‌ای که از آن ناشی شده محدود می‌ماند. در انتخاب یک گروه نمونه، باید تلاش شود تا به یک نقطه عطف شاخص در گروه که آزمون بدان منظور طرح و تنظیم شده است دست یافت. گذشته از این، گروه نمونه بایست بقدر کافی بزرگ باشد که نمرات پایائی را به‌نمایش بگذارد؛ معمولاً گروه نمونه‌ای با حدود ۲۰۰ نفر توصیه می‌شود.

۲-۵-۷ تفسیر ملاکی

گاهی ممکن است ارزیابی صرفاً بخاطر اینکه معلوم شود آیا آزمون شونده به اهداف معینی نائل آمده است یا خیر انجام میشود. منظور از آن تفاوت‌گذاری بین افراد نیست. تفسیر ملاکی آنگونه که خواننده میشود، بر این امر تأکید دارد که آزمون‌دهنده با آنچه که میداند چه میتواند بکند، نه اینکه با دیگران چگونه مقایسه میشود. بدین ترتیب بهنگام تألیف آزمون‌های ملاکی اساساً زمینه مهارتی که قرار است سنجیده شود باید بروشنی تعریف شده باشد. این زمینه برگزیده سپس به بخش‌های کوچک‌تری تقسیم میشود، هر یک با نام معینی مثل تشخیص درست عبارات موصولی شرطی که آزمون دهنده باید انجام دهد. بعد از تعیین اهداف سئوالات آزمون آماده میشود. مثلاً تصور کنید خواسته باشیم مهارت فردی را در درک مطلب خواستار باشیم. ما میتوانیم عبارات بی‌شماری از مطالب مختلف در آزمون جمع‌آوری کنیم که هر کدام شامل تعدادی سئوالات درک مطلب باشد. در واقع، این امر شدنی نیست. زیرا که نیازمند زمان زیادی برای برگزاری است که مافوق تحمل و طاقت هم مسئولان و هم داوطلبان است. بدین لحاظ ما میتوانیم به تنها ۴۰ سئوال درک مطلب بر مبنای هفت پاراگراف که به اختیار از یک آزمون بلند استخراج شده است رضایت دهیم. چنانچه فرد به ۲۸ سئوال جواب درست داده باشد، معلوم است که قادر به درک ۷۰ درصد مطالب می‌باشد. در هر حال، بعداً ممکن است دریابیم که این‌گونه توانمندی افراد در قرائت شرایط ما را برآورده نمیسازد؛ فقط آنهایی که ۳۲ نمره به بالا گرفته بودند خوانندگان توانائی بودند. در نتیجه ممکن است به افزایش سطح معیار تا ۸۰ درصد درک مطلب تصمیم بگیریم.

روش ملاکی ویژگی‌هایی خاصی دارد. معمولاً، یک داوطلب تنها وقتی به تمامی سئوالات یا تعداد معینی از آنها پاسخ صحیح می‌دهد، قبول می‌شود. چون هدف آزمون اینست که ببینیم آیا آزمون‌شونده به مرحله معینی از تسلط (مثلاً به نوعی مهارت) رسیده است یا خیر. والا نمره بالاتر داشتن چندان فرقی ایجاد نمی‌کند. البته نمره بالاتر ممکن است وقتی فرق کند که داوطلبان آموزش تقویتی ببینند. هنگامیکه آزمون‌شونده ۹۰ درصد نمره کل آزمون یعنی همان استاندارد قبلی را کسب کرده باشد، در حقیقت امتحان را قبول شده است، آنگاه برای آنچه که موفق به پاسخ نشده است آموزش تقویتی می‌بیند. در صورتی که سئوالی یا سئوالاتی توسط شمار بزرگی از آزمون‌شوندگان بدون پاسخ مانده باشد، آن سئوال از آزمون حذف می‌شود، آنطور که در آزمون‌های متعارف مرسوم است. در عمل اگر آزمونی بمنظور دوره خاص تنظیم شده باشد، مطالب آموزشی مورد تجدیدنظر قرار می‌گیرد یا کلاس تقویتی گذاشته می‌شود. انتخاب آزمون‌هایی که، منظور از آن‌ها تهیه اطلاعات پیرامون قبول شدن یا نشدن در برنامه‌ای خاص است، نشان از امتحانات تفسیر معیاری دارد مثلاً، متقاضی گواهی‌نامه رانندگی آیا مهارت در رانندگی را بگونه‌ای رضایت‌آمیز کسب نموده است یا نه. آزمون‌های تخصصی ممکن است به عرصه‌ی آزمون‌های معیاری تعلق پیدا کنند. مثلاً مدارس گرجوئ و وابسته به دانشگاه‌های آمریکا دانشجویانی که دوره زبان تافل را گذرانده‌اند و حداقل نمره ۵۵۰ را در دوره تافل کسب نموده‌اند می‌پذیرد. این روش تفسیر نمرات در شرایط امتحانات ملاکی محقق می‌گردد. هنگامیکه آزمونی طی مراحل مختلف زمانی برگزار می‌شود، اطلاعات گردآوری شده می‌توانند بمنظور پیش‌بینی عملکرد آینده آزمون‌شوندگان در یک دوره یا برنامه تحصیلی بکار آیند.

برای تهیه یک چنین جدول مورد انتظاری به نمرات آزمون شمار بزرگی از افراد و نتایج کارهایشان در برنامه یا کورس تحصیلی‌شان نیاز هست. جدول فرضی انتظارات در شکل ۳-۷ آمده است.

جدول ۳-۷

High school GPA	Number of cases	Percentage Receiving Critierion Grade college GPA			
		Dropped out	C	B	A
18.20.0	200			6	94
10.17.9	169		4	80	16
14.15.9	201	9	70	19	2
13.13.9	61	20	70	10	
Below 13	11	80	18	2	

ستون ۱ میانگین نمره (GPA) دیپلم دبیرستان افراد را نشان می‌دهد. ستون ۲ اندازه هر گروه را مشخص می‌سازد. مندرجات ستون ۳ نشانگر درصدی افرادی است که از لحاظ تحصیل در کالج مردود شده‌اند و افرادی که میانگین نمرات کل کالج را کسب کرده‌اند بترتیب در ستون C , B , A قرار دارند. بدین ترتیب از میان ۲۰۰ دانشجوی با میانگین نمره ۱۸ به بالا ۶۰ درصد با معدل B و ۹۴ درصد با

نمره A فارغ‌التحصیل شده‌اند. در طرف دیگر، از میان ۱۱ نفر، معدل زیر ۱۳، ۸ درصد مردود شده‌اند. ۱۸ درصد با معدل C قبول شده‌اند و تنها ۲ درصد معدل‌شان B بوده است.

در تفسیر یک چنین جدولی اهمیت نکات زیر خاطرنشان می‌شود. نخست، اندازه گروه از اهمیت برخوردار است. ستون ۲ نشان می‌دهد که درصد آخرین ردیف تنها بر ۱۱ نفر متکی است؛ یک چنین ارقامی بشدت در معرض نوسان قرار دارند. دوم اینکه اطلاعات جدول با اتکاء به گروه خاص تنظیم شده است. چنانچه داده‌هایی از یک نمودار متفاوت در اختیار می‌داشتیم، نمرات وضعیت دیگری داشتند. بدیهی است. جدول رخدادهای حاکی از تفسیر معیاری نمرات آزمون و نیز سندی بر اعتبار پیش‌بینی آزمون می‌باشد. لازم به تذکر است که تفسیر متعارف و معیاری هر دو مضرند.

معهدا با توجه به تصمیمی که ما می‌گیریم، یکی از دو روش تفسیر نمرات مسلماً از سایرین مفیدتر خواهد بود. در واقع دو روش ناقض همدیگر نیستند. برای هر یک جایگاهی وجود دارد. سیستم آموزشی فعلی بطور معمول با آزمون‌های روش متعارف سازگاری دارد اما بهر حال چنانچه تسلط در مهارت‌های رفتاری حق تقدّم یابند (برتر به حساب آیند) آنگاه تفسیر معیاری ترجیح داده خواهد شد. با تأکید بر آزمون‌های شفاهی - گفتاری، تفسیر معیاری در حال حاضر و بخصوص تست‌های مهارت طرفدار بیشتری دارند.

فعالیت‌ها

نکات زیر را مورد بحث قرار دهید:

- ۱- به چه دلیل تغییر در مورد روانشناسی بر آزمون‌سازی تأثیر می‌گذارد.
- ۲- چگونه یک شخص بر اساس محتوای یک آزمون، نوع آن را به عنوان مجزا شکل، یکپارچه، عملی یا کارکردی تعیین می‌کند؟ برای هر کدام مثالی بزنید.
- ۳- آیا ممکن است آزمون گرامری چند گزینه‌ای که یکپارچه تر از مجزا شکل است طرح کنیم؟
- ۴- تفاوت میان آزمون‌های یکپارچه و عملی چیست؟
- ۵- آیا می‌توانید تفاوتی میان آزمون‌های عملی و کارکردی پیدا کنید؟

آزمون واژگان

۸-۱ مقدمه

هدف از آزمون لغات، سنجش معلومات واژگانی افراد می‌باشد. در تهیه و تنظیم آزمون واژگان، نخستین کار عبارت است از گزینش کلمات آزمون. در آزمون‌های پیشرفتی مؤلفین آزمون سئوالات واژگان را از همان مطالب درسی ضمن دوره انتخاب میکنند. با این حال بهنگام آزمون مهارت (پرافیشنی) زبان کار مؤلف سخت‌تر میشود. او باید تصمیم بگیرد که آیا اندوخته لغات فعال یا غیرفعال داوطلب را آنهم تا چه میزان بیازماید، و اینکه گزیده کلمات باید با زبان گفتاری و یا نوشتاری سازگاری داشته باشد یا خیر. بطور کلی باید خاطر نشان سازیم، که در دوره مقدماتی سئوالات باید از میان کلمات معمولی (دم دست) انتخاب شوند. مثل اسامی اشیاء توی کلاس و یا متعلق به افراد. در سطح متوسط، کلماتی که در صحبت کردن و مکالمات شفاهی لازم می‌ایند باید شامل شوند. در مرحله پیشرفته کلمات باید از میان واژگان زبان نوشتاری برگزیده شوند. البته کلمات معنائی، اسامی، افعال، صفات، قیود در آزمون واژگان می‌ایند. لغات کارکردی، حروف تعریف، ضمائر اشاره و حروف اضافه ربط، ضمائر و افعال معین در آزمون گرامر زبان در فصل بعد مورد بحث قرار خواهند گرفت.

در مورد انتخاب بین کلمات غیرفعال و فعال، تفاوت قائل شدن تقریباً کار راحتی است. اولی دلالت بر کلماتی میکند که داوطلب در نوشتار یا گفتار آنها را می‌یابد. اما خود در نوشتن یا حرف زدن به کار نمی‌برد. و دومی با همانهایی هستند که داوطلب در گفتار و نوشتار بر آنها اشراف کامل دارد.

پس از آنکه پیرامون ملاحظات فوق تصمیم گرفته شد، طراح آزمون باید حساب توزیع فراوانی، اندازه و قابلیت استفاده در آزمون را مورد توجه قرار دهد. فهرست‌های کلمات برای مشاوره مثل

فهرست کلمات انگلیسی برای استفاده عمومی (غرب، ۱۹۵۳) بر مبنای فراوانی سئوالات واژگان می‌تواند روشنگر و مفید باشد. مطمئناً، چنین فهرستهایی بدلائل متعددی عموماً دارای موارد استفاده محدودی هستند، که عمده‌ترین‌شان اینهاست. نخست اینکه اغلب مصطلح نیستند. دوم اینکه براساس داده‌های زبان نوشتاری متکی هستند. و بالاخره، وجوه سختی کلمات را بویژه در ارتباط با زبان اصلی داوطلب نشان نمیدهند. بدین ترتیب برای جبران چنین نارسائی‌ای، لازم می‌آید تا طراح آزمون به نظر خود یا نظریات تجربی همکاران در رابطه با اهداف آزمون متکی بماند.

وظیفه دوم طراح آزمون در تهیه و تنظیم آزمون واژگان، تعیین فرم سئوالات می‌باشد. شکل ایدال برای کلمه‌ای که قرار است آزموده شود اینست که زیر آن در متن خط کشیده و چهار معنی برای آن در نظر گرفته شود.

چهار گزینه معمولاً عباراتی برای همان کلمه موردنظر در متن می‌باشد.

Example 1

After discussing the matter for two hours , the committee adjourned without having any decision,

a) finished

c) took off

b) continued

d) broke off

این روش خوبی برای آزمودن میزان درک داوطلب دایر بر معنی واژگان یا عبارات است. این تکنیک با وجود این سه عیب دارد. اول اینکه امر آزمون را تنها به یک کلمه وابسته می‌سازد. دوم، اینکه آیتمهای واژگان گاهی به چهار گزینه منطقی منتهی نمی‌شوند. و سوم، موجب میشود که داوطلب تمامی جمله را فراموش کند و فقط به معنی کلمه توجه داشته باشد. بدین ترتیب انواع این روش‌ها که از معایب برشمرده در فوق بری میمانند اغلب جزء آزمون‌های واژگان جای میگیرند. این نوع سؤال طوری تنظیم میشود که جای کلمه موردنظر در متن خالی بماند. داوطلب جمله را میخواند و سپس مناسبترین کلمه که جمله را کامل می‌سازد، برمیگزیند.

مثال ۲

After discussing the matter for two hours, the committee without having reached

a- debated

c- collapsed

b- grouped

d- adjourned

این تکنیک در عین حال با صرفه عمل میکند، چرا که در یک زمان چهار کلمه را همزمان بمدد معلومات داوطلب محک میزند.

تست این گونه که معمولاً در میان آزمون‌های واژگان خیلی شهرت دارد در اینجا مورد تأیید نیست. آزمون واژگان نمونه‌ای استاندارد است.

Example 3

"Adjourn" means

a- Cause to turn aside

c- come into use again

b- break into pieces

d- break off for a time

Example 4**“Break off for a time” means**

a- collapse

b- adjourn

c- deflect

d- prostrate

این خلاصه‌ترین نوع سؤال واژگان است. اما بدلیل تأثیرات جانبی آن، در اینجا توصیه نمی‌شود. یک چنین شکلی داوطلب را به حفظ کردن فهرست کلمات وامیدارد که بلحاظ آموزشی قابل توصیه نیست. یک متن مناسب و با معنی بهر حال فرصتی در اختیار آزمون شونده میگذارد تا بتواند معنی کلمه موردنظر را حدس بزند. گذشته از همه آزمون نیز بایستی اغلب فرصتی برای یادگیری فراهم آورد. شکل استاندارد این نیاز را برطرف نمیسازد. بهمین دلایل نمونه سؤال زیر مناسب‌تر برای آزمون واژگان است.

این گونه آیتم شامل متنی مرتبط با چند لغت معنایی حذف شده است. و داوطلب جاهای خالی را با کلمات مناسب از خود پر می‌سازد.

Example 5

Good drivers always obey the policeman. They do what he tells them. But Sometimes a bad driver the policeman, then he takes out his notebook. He writes down the number of the Later , adjudge may make the bad driver pay some money.

۲-۸ رهنمودی برای تألیف سئوالات واژگان

بهنگام تهیه و تنظیم آیتمهای واژگان، رهنمودهای زیر را باید بخاطر سپرد.

۱- متن بایستی بقدر کافی بدور از ابهام باشد تا معنایی روشن برای داوطلبان داشته باشد. یک متن گفتگو گونه کوتاه از یک متن کوتاه غیر آن مناسب‌تر است. بویژه برای افرادی که مهارت کمتری دارند. مثلاً در مورد واژه borrow در عبارتی شبیه «Behnam borrowed John's book» ضعیف عمل میکند زیرا که تنها نشان میدهد که borrow یک فعل است. در حالیکه، در عبارت دیگری شبیه عبارت زیر چهارچوب بهتری برای این کلمه ارائه میدهد.

“Did Behnam buy another book?”**“No, he John's”**

a- borrowed

b- sold

c- lent

d- returned

۲- در حین تنظیم سئوال باید دقت کرد تا ساختار گرامری یا نکته‌ای اشتباه و مشکل ساز که بنظر داوطلب سخت بیاید و مانع از درک وی گردد پیش نیاید. در جمله زیر با اینکه کلمه آزمون در خود متن جای دارد، اما در هر حال جمله ضعیفی است زیرا که درک متن خود جمله برای داوطلب در این مرحله از مهارت زبانی خیلی دشوار است.

Being unfortunate to have been bereaved of his belongings, Reza Lucky Johu's book.

a- sold

b- lent

c- borrowed

d- returned

در همان راستا، سؤال زیر نیز ضعیف است، چرا که یافتن جواب درست نیازمند آنست که داوطلب معلوماتی و رای تسلط بر این گونه آیتمهای واژگان داشته باشد:

We can raise sugar cane in khusestan because cane needs a climate

a- dry

c- warm

b- wet

d- cold

۳- در صورتی که سؤال از نوع عبارت گونه (مثال ۱) باشد، گزینه‌ها باید ساده‌تر از خود کلمه موردنظر باشند. بعلاوه بلحاظ گرامری باید مشابه شکل گرامری کلمه زیر خطدار باشد. یعنی اینکه، گزینه باید بسهولت برای داوطلب قابل درک باشد و بلحاظ نحوی جایگزین قابل قبولی برای کلمه موردنظر باشد. نمونه‌های زیر حاکی از یک سؤال نارساست، زیرا که نخستین نیازی که هم‌اکنون برشمردیم در ترکیب این سؤال رعایت نشده است.

The child was frightened of being left alone in the dark room

a- annoyed

c- terrified

b- ashamed

d- dismayed

این سؤال چهارگزینه دارد، که همه در یک سطح از دشواری قرار دارند، با این حال یقیناً بهمان اندازه کلمه مورد اشاره در متن سخت و مشکلند. ترکیب کلمه سوّمی از کلمه خود متن غیرمصلحت‌تر می‌باشد. بنابراین، اگر داوطلب در پاسخ دادن به این سؤال توفیق نداشته باشد، نمیتوان با اطمینان گفت که وی کلمه "frighten" را نمیدانسته، بلکه در حقیقت ممکن است معنی واژه "Terrify" برایش دشوار آمده باشد.

سؤال زیر نشان از برخی اصلاحات در جمله فوق‌الذکر دارد، گرچه همچنان دچار نقص دیگری است: در میان چهار گزینه، تنها یکی بلحاظ ترکیب گرامری جایگزین کلمه متن می‌باشد. در نتیجه، توانائی پاسخ صحیح به این سؤال الزاماً ربطی به دانستن معنی کلمه درون جمله ندارد، تنها میشود صرفاً با اتکاء به معلومات گرامری پاسخ درست سؤال را یافت. وقتی موردی بدین‌گونه پیش بیاید، چنین سؤال‌ی درک معنی کلمه "frighten" را از جانب داوطلب خواهد سنجید:

The child was frightened of being left alone in the dark room.

a- made somewhat angry

c- attacked suddenly

b- feeling shame

d- filled with fear

همین‌طور سؤال زیر جنبه دیگری از نارسائی را نشان میدهد. در این آیتم، تنها یکی از چهار گزینه، اگر فرض کنیم که کلمه زیر خطدار اصلاً در میان نباشد در جمله معنی میدهد. بازپس این سؤال هم معلومات داوطلب را با توجه به واژه موردنظر در عبارت نشان می‌دهد.

The detective is investigating a hidious crime

a- a torrential

c- an insensible

b- a detestable

d- an immaculate

۴- چنانچه سؤال از نوع تکمیلی باشد (مثال ۲)، واژه‌های انحرافی و کلمه مورد آزمون از لحاظ دشوار بودن در یک سطح قرار دارند. علاوه بر این، انحرافی‌ها بایست بلحاظ نحوی و بن کلمه قابل قبول باشند. مثلاً، سؤال‌ی مثل نمونه زیر سؤال ضعیف مینماید. چرا که دو گزینه نخست از دو تایی دیگر

آسان ترند. داوطلب بسهولة میتواند دوتای نخست را غلط بداند و با توجیه غلط به پاسخ درست نزدیک شود.

Ahmad says he would write an English course book if he could find a to deal with the less interesting parts.

- | | |
|-----------|-----------------|
| a- reader | c- lithographer |
| b- Tailor | d- collaborator |

و سئوالی نظیر این:

I got up so late that I had time for breakfast.

- | | |
|-----------|-----------|
| a- dull | c- hardly |
| b- rarity | d- yet |

این جمله شامل انحرافی‌هایی است که ضعیف عمل میکنند. زیرا که داوطلب میتواند بسهولة دوتای اولی را بدلیل شکل نحوی نامناسب از دور خارج سازد.

۵- گزینه‌ها تا حد امکان بایستی به موضوع یا زمینه‌ای عموماً یکسان مربوط باشند.

She came to the party in a dress

- | | |
|------------|------------|
| a- capable | c- Lunalic |
| b- sincere | d- hideous |

چنانچه، برای مثال، داوطلب بمحض مواجهه با چنین سئوالی مطمئناً سه گزینه نخست را کنار خواهد گذاشت چرا که اصلاً با لباس پوشیدن جور در نمی‌آید. آیتم زیرین در هر صورت نیاز بیشتر به تسلط بر سئوالات واژگان دارد.

She came to the party in a hideous

- | | |
|-----------|-----------------------|
| a) a nice | c) a very large dress |
| b) an old | a very ugly |

ع گزینه‌ها بایستی تقریباً به یک اندازه باشند. هیچکدام آنها نبایستی بلندتر یا کوتاه‌تر از دیگری باشد.

Ahmad says he would write an English course book if he could find To deal with the less interesting parts.

- | | |
|-------------|-------------------|
| a- a vet | c- an ace |
| b- an idiot | d- a collaborator |

گزینه آخری سئوال بالا، دستکم بلحاظ ظاهر با آنهای دیگر متفاوت است.

گمان می‌رود که شمار زیادی از داوطلبان گزینه (d) را بنابه دلایلی انتخاب خواهند کرد. برخی از کارشناسان معتقدند که یک چنین نقضی ناشی از وسواس مؤلفان آزمون می‌باشد که بعد میکوشند تا درست بودن این گزینه را هر طور شده تأیید کنند.

- | | |
|------------------|--------------|
| a- an arbitrator | c- a partner |
| b- a proofreader | d- an editor |

فعالیت‌ها

1. What are the advantages and/or limitations of each item type discussed in this Chapter?
2. What do you think of the following types of items?
A. Which word is associated with the word vehicle?

- a- bus
b- driver
c- asphalt
d- journey

B. Which of the following words best represents pork, beef, chicken, and fish?

- a- animals
b- meat
c- beverages
d- flesh

C. Ask a student what color his/her book is.

D. Ask a student to give you the green book on the desk.

E. Fill in the missing letters in the incomplete words:

You cannot see the wind, but you know when the w.....is blowing. You can.....1 it against your f..... You can see the l.....on trees moving as the.....d blows against them.

F. Fill in the missing words:

Wind is air that is moving. The faster the.....moves, the stronger the.....is. A gentle, slow-moving.....is called a A hurricane is a that is moving very, very

3. In the Teachers's word book of 30,000 words (Thorndike & Lorge 1944) the words "microbe" and "khaki" are ranked in frequency with "mullusc" and "reticent". Which one(s) would you include in your proficiency test for Persian EFL/ESL learners? Why?

4. Each of the following items has some defects. Point out the major weakness (es) in each and then modify each item to the extent possible:

A. The first of the week is.....

- a- Monday
b- Friday
c- Sunday
d- Saturday

B. Ahmad did not go to school today because he was ill.

- a- weak
b- sick
c- feeble
d- injured

C. "I don't like its appearance. It looks rather square and large. I think it's far from beautiful; in fact, I think it's....."

- a- graceful
b- afterward
c- clumsy
d- reasoned

D. After discussing the matter for two hours, the committee.....without having reached any decision.

- a- ended
b- deferred
c- postponed
d- adjourned

E. Columbus.....America.

- a- invented
b- explored
c- created
d- discovered

F. Before buying the house, I pondered a lot on its consequences.

- a- meditated
b- surmised
c- conceived
d- estimated

G. She came to the party in a hideous dress.

- a- color picture
c- delicious dessert

b- short sleep

d- long journey

5. Using the following list, write one item of the paraphrase type (Example 1) and another of the completion type (Example 2):

official

include

achieve

possess

collect

efficient

drowsy

contain

فصل نهم

آزمون گرامری

۱-۹ مقدمه

به سه دلیل گرامر فراگیرترین مؤلفه زبان به حساب می‌آید. نخست، اینکه در تمامی مهارت‌های زبانی دخالت دارد: دوم از سایر مؤلفه‌ها برای امتحان گرفتن راحت‌تر است: و سوم بیشتر کارشناسان در اینکه چه نکاتی باید در آزمون گرامر گنجانده شود توافق نظر دارند.

آزمون گرامر میکوشد تا معلومات داوطلب را در زمینه شکل و انتظام درست کلمات، عبارات، و جملات بسنجد. آزمون گرامری افرادی که زبان مادریشان انگلیسی است بر گرامر نوشتاری تمرکز دارد، مباحث گرامری زبان آموزان خارجی، از طرف دیگر، با الگوهای گرامری متناسب برای مکالمه سروکار دارد. تنها بخاطر داوطلبان مرحله پیشرفته زبان، هدف از آزمون، اندازه‌گیری معلومات سیستم گرامری بمنظور مکالمات رسمی است.

الگوی معمولاً مناسب معلومات گرامر زبان شامل یک گفتگوی کوتاه دو طرفه است دو جمله‌ای با جای خالی بهمراه چهار گزینه که یکی از آنها مکالمه را کامل می‌سازد.

“Did Ali go shopping alone?” “No, but he brought the tie”

a- to himself

c- for himself

b- by himself

d- with himself

در مواقعی که محدودیت جا اجازه یک چنین متن آزادی را نمی‌دهد، یک شکل کوتاه‌تر میتواند بکار آید. این شکل معمولاً یک جمله نیمه تمام است که چهار گزینه دارد، یکی از آنها جمله را کامل می‌سازد.

مثال ۲

This box is heavier That one

a- of

c- than

b- to

d- from

نوع سوم سنوآل گرامر نوع درهم ریخته است، و داوطلب باید این اجزا را مرتب کند تا جمله گرامری شکل بگیرد.

مثال ۳

This box is one / heavier / that / than**A B C D**

a. ABCD

b. BDCA

b. ACDB

d. CBDA

در سنوآل فوق از داوطلب انتظار می‌رود تا این نکات را بترتیب BDCA مرتب کند.

تکنیک درهم ریخته مناسب حال زبان آموزان نوجوانی است که معمولاً علاقمند به کارهائی با راه حل پازلی هستند. این روش بهر صورت، کاربرد محدودی دارد علاوه بر آن، بلحاظ شکل این آزمون، نکات زیادی را نمیشود بدین وسیله آزمود. بویژه در مورد زبان آموزان مرحله پس از مقدماتی و پیشرفته.

در تهیه و تنظیم سنوآلات گرامری باید دقت شود تا هر نکته‌ای در متن مناسب جا بیافتد. متن بایستی تا حد ممکن مطابق با شرایط طبیعی و معنی‌دار باشد. بایستی کوتاه، روشن، و سر راست بدون هرگونه کلمه آشنائی برای افراد داوطلب باشد. اصل ثابتی در مورد اندازه متن وجود ندارد، گاهی یک متن چند کلمه‌ای کفایت میکند. در حالیکه در موارد دیگر یک متن ۲۰ کلمه‌ای هم ممکن است کم باشد.

استفاده از انحرافی‌هائی که موجه باشند نیز اهمیت دارد. معلمان باتجربه معمولاً معتقدند که انحرافی‌ها برای هر متن معینی مناسب‌ند. معلمان بی‌تجربه تنها به اشتباهاتی که دانش‌آموزان در تمرینات کلاسی، انشاء‌ی آزاد و پاسخهای که به سنوآلات جا خالی داده میشود نظر دارند. علاوه بر آن انحرافی‌ها بر مبنای تحلیل‌های افتراقی میتوانند انتخاب گردند، این در صورتی است که زبان داوطلبان یکی باشد. گذشته از این، سنوآلات گرامری به شکل سنوآلات تکمیل جملات، جواب کوتاه، و سنوآلات نیمه تمام مطرح هستند. این‌ها روش‌های مناسبی هستند، بویژه بمنظور تمرینات کلاسی، یا وقتی شمار داوطلبان خیلی زیاد است.

مثال ۴

This box is heavier That one.

مثال ۵

Is This box as heavy as that one?**No, it's than that one.**

یک شکل مناسب دیگر پاراگراف یک مکالمه مرتبط با چند تا کلمه کارکردی حذف شده است که آزمون شونده باید جای خالی را پر سازد. این جای خالی ممکن است برای یک دسته نکات گرامری باشد (مثلاً حروف تعریف) یا برای نکات کارکردی مختلف:

مثال ۶

Good drivers always obey the policeman. They do what he tells them. But sometimes.....bad driver disobeys.....policeman. then.....policeman takes his notebook. He writes down.....number of.....car.

Later a judge may make the bad driver pay some money.

این تکنیک متن پر کردن روش جا خالی (Cloze) نام دارد. ما درباره شیوه تهیه و تنظیم یک آزمون جا خالی را در فصل بعد بحث خواهیم کرد.

۲-۹ رهنمودی برای تهیه و تنظیم سوالات گرامر

پس از آنکه نکات گرامری گزیده شدند، مؤلف به نوشتن سوالات واقعی آزمون میپردازد. در این زمینه رهنمود زیر توصیه میشود.

۱- بن سؤال بایستی نشانه یک مکالمه واقعی در یک برخورد اتفاقی باشد. متنی نظیر،

“The professor demanded that his students be on time to every class.”

خیلی رسمی و کتابی است و اگر Asked به demanded تغییر کند برای یک متن با سوالات گرامری مناسب تر خواهد بود. و نیز:

“Are you going home?”

“Yes I am, I'm home.”

طنین یک فعالیت مکانیکی کلاسی را دارد. استفاده زبانی در شرایط واقعی مطمئناً امکان ندارد.

۲- هر سؤالی بایستی تنها یک پاسخ قابل قبول داشته باشد. آیتمها نبایستی شامل انحرافی‌هایی باشد که در یک ناحیه یا در گونه‌ای زبان مصطلح هستند و نه در جاهای دیگر. مثلاً، از سؤالی نظیر سؤال زیر بایستی احتراز کرد، چرا که انحرافی اولی مناسب زبان انگلیسی بریتانیایی است و نه انگلیسی امریکائی. این مستلزم آنست که تمامی انحرافی‌ها، کلاً غلط باشند.

“Why are you ringing the door bell?”

“I my key with me.”

a- haven't

c- don't have

b- don't

d- not have

۳- بن بایستی متن مناسب را فراهم سازد. و گزینه‌ها باید کوتاه و مربوط باشند.

Example

“what did your father ask you?”

“He asked me younger brother to school “.

a- if I had taken my

b- If I have taken my

c- Did you take your

d- Did I take my

- a. I asked Mary had she bought what yesterday.

b. I asked Mary what had she bought what yesterday.

c. I asked Mary she had wht bought yesterday.

d. I asked Mary what she has bought yesterday.

C. Rewrite the following in another way, beginning with the given word:

On the way to school I met on of my old friends.

Going.....

D. Ask a question with "who"/.

E. Match the question/expression in Column I with possible answers in Column II;

I	II
1. How is Tom?	a. You're welcome.
2. Mary I smoke here?	b. Much better today.
3. Who is at the door?	c. Certainly.
4. Thank you.	d. Tom is.
	e. They came at 6.

F. Fill in theblank with a suitable connectors:

It was raining heavily, we decided to go out.

G. Change the following into indirect speech:

Mary asked, "What time is it?"

.....

3. Each of the following items has some defects. Points out the major weakness (es) in each and then modify each item to the extent possible:

1. "..... did you meet Ali?"

"at 5:30 p.m."

a- Where

b- Why

c- When

d- What

2. Whose book.....over there?

a- is that

b- is this

c- are those

d- are these

3. If I see Ali, I.....him your messages.

a- give

b- gave

c- will give

d- would give

4. "What is Ali's profitable business?"

a- Wrote

b-Writes

c- Writing

d- Written

5. "What was their reaction when they saw their names on that list?"

a- As a result

b- Therefore

c- Consequently

d- However

6. "Why did they give John a prize?"

"Well, he got the.....grade point average."

a- top

b- above

c- higher

d- highest

7. Which of the following sentences is not correct?

a- Who asked you to buy this pen?

- b- The bird that I saw in the garden had a strange name.
- c- My brother which is in Shiraz wants to study medicine.
- d- The branch the gardener removed from the tree was dead.

4. Write one item testing relative clause construction and another testing the formation of comparisons in English. Use Examples 1 as a model.

فصل دهم

آزمون طرز تلفظ

۱-۱ مقدمه

برای تسلط بر زبان، لازم است تا بدانیم که حروف دارای چه صدائی هستند و برحسب قواعد آن زبان طبق چه الگویی عمل میکنند. اشراف کامل یا چیزی نزدیک بدان بر سیستم صوتی زبان خارجی که بر طبق اصول آوایی قابل قبول باشد. یک هدف دلخواه هست، اما جنبه مهم فرایند یادگیری نیست. در واقع، همانهایی که انگلیسی زبان مادریشان می باشد تفاوت های فونتیکی را آنطور که باید رعایت نمیکنند. آنها آنچنان با برخی الگوها و عبارات آشنائی دارند که بدون توجه چندانی به تک تک کلمات مطلبی را درک کنند. در ضمن، برای دو دسته که زبان انگلیسی زبانی غیرمادری است بدون آنکه تلفظ درستی داشته باشند، امکان مکالمه ب زبان انگلیسی وجود دارد. در نتیجه، معیار صحت آزمون های عمومی تلفظ زبان بایستی بر قابل درک بودن نهاده شود. بدیهی است که رعایت ملاک های عالی تلفظ صحیح بستگی به هدف یادگیری زبان دارد.

آزمون های تلفظ، در نتیجه، میکوشند تا توانائی داوطلب را بمنظور تشخیص وادای اصوات، الگوی تکیه ها (استرس) و لحن صدا در زبان بروش قابل قبول واجی بیازمایند. بویژه، آزمون های تلفظ از مبتدیان و داوطلبان مرحله متوسط شروع می شود. در رابطه با داوطلبان مرحله متوسط به بالا و پیشرفته بیشتر سنجش درک شنیداری اهمیت دارد. مباحث زیر حاکی از متداولترین روش های سنجش مهارت تشخیص وادای اصوات و الگوی های ترکیب آنها در زبان انگلیسی است.

۲-۱۰ تشخیص

آزمون‌های تشخیص اصوات، تکیه‌ها، و طنین صدا بخوبی می‌تواند از طریق این گونه سؤالات نظیر چندگزینه‌ای و صحیح -- غلط که داوطلب را از زحمت خواندن برهاند و فرصت دهد تا به محرک‌های شنیداری دقت کند، صورت می‌گیرد.

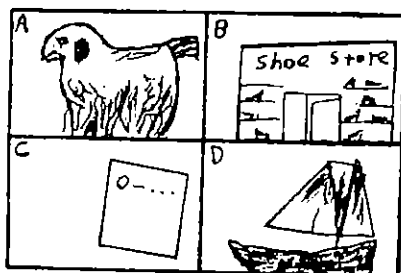
۲-۱۰-۱ تمایز صوتی

صوتها از طریق تصاویر یا بدون آن قابل آزمودن می‌باشند. کاربر آزمون مثلاً کلمه یا عبارت کوتاهی را می‌خواند و داوطلب از میان تصاویر یکی را که مطابق با محرک صوتی است برمی‌گزیند.

مثال ۱

Examinee hears
That's a ship

Examinee reads
A B C D



سؤالات تصویری بویژه برای مبتدیان و کودکان مفیدند: اینها وسیله خوب مناسبی برای رساندن پیام به داوطلبان بدون هرگونه نیازی به خواندن و نوشتن هستند و ی کودکان هم جالب و دلچسب است. البته عکس‌ها هم می‌توانند در صورتی که خیلی روشن تصویر نشده باشند مبهم جلوه کنند. گاهی تصاویر بفوریت توجه را به شیئی یا عملی جلب نمی‌کنند. در واقع، پاره‌ای کلمات قابلیت تصویر شدن ندارند. گذشته از این تهیه عکس‌های مناسب گران و وقت‌گیر است. در یک چنین وضعیتی، تست صوت به تنهایی صورت می‌گیرد. در این گونه سؤالات کاربر کلمه‌ای را می‌خواند و داوطلب بدقت می‌شوند و سپس کلمه‌ای را که شنیده است بدرستی از میان چهار گزینه یا بیشتر در روی پاسخ نام انتخاب می‌کند.

Example 2

examinee hears
Sheep

examinee reads

- A) shop
- B) ship
- C) sheep
- D) shape

نمونه‌های مختلفی از این گونه شامل محرک‌های شنیداری است که عبارتند از سه یا چند کلمه و داوطلب باید یکی را که با آن‌های دیگر فرق می‌کند تشخیص دهد.

Example 3

examinee hears
Cat , it , eat

examinee reads

A B C

از آنجا که تهیه و اجرای یک چنین سئوالاتی آسان است، کارکردهای مفیدی می‌تواند داشته باشد. با این وجود این نوع سئوالات نه معنایی دارند و نه با هیچ نوع فعالیتی در زندگی واقعی مطابقت می‌کنند. نمونه آزمونی که این محدودیت را ندارد دیکته می‌باشد.

در تست دیکته، کاربر متنی را برای داوطلبان می‌خواند و از ایشان می‌خواهد تا آنچه واحد می‌شنوند بنویسند. البته، دلیل آنکه داوطلبان برای آنکه ادراک صوتیشان را نشان دهند ناچارند تا آنچه را که می‌شنوند بنویسند، پس تست دیکته هم نارسائی‌هایی دربردارد. توضیح مشروح درباره دیکته نویس و آنچه که غیر از درک صوتی می‌سنجد در فصول بعد خواهد آمد.

۲-۱۰ تشخیص تکیه

تکیه طبق رسم دیرینه بطور مجزا مورد آزمون قرار می‌گیرد. ممتحن کلمات دو بخشی یا بیشتر را می‌خواند و از داوطلب می‌خواهد تا بخش پرفشار را مشخص کند.

Example 4

examinee hears
English
Laboratory

examinee reads

1 2
1 2 3 4 5

در همین راستا ممتحن جمله‌ای را می‌خواند و از داوطلب می‌خواهد تا کلمه‌ای را که تکیه می‌پذیرد علامت زند: و بدین ترتیب تست استرس (فشار - تکیه) جمله صورت می‌پذیرد.

Example 5

examinee hears
The Teacher will
Teach lesson two
Or the teacher will
Teach lesson two

examinee reads

1 2 3 4 5
1 2 3 4 5

در هر صورت، سنجش میزان تشخیص داوطلب از بابت الگوهای استرس در متن با جابه‌جایی افتراقی استرس در آن به بهترین وجه ممکن می‌گردد. جمله‌ای مثل "Good boys and girls study hard." می‌تواند معانی مختلفی برحسب آنکه روش کدام کلمه تکیه را قرار داده باشیم، داشته باشد.

برای مثال، چنانچه فشار اصلی روی بر "boys" باشد جمله بدین معنی خواهد بود "both good boys and good girls study hard" در هر صورت اگر فشار اصلی روی "and" باشد، معنای آن این است که تنها پسرهای خوب و تمامی دختران سخت مطالعه میکنند

Example 6

examinee hears

Good boys and
Girls study hard,
Good boys and
Girls study hard.

examinee reads

Same different

Example 7

examinee hears

Good boys and
Girls study hard.

examinee reads

- A. Boys and girls Study hard.
- B. All boys and good girls Study hard.
- C. Only good boys and all Girls study hard.
- D. Both good boys and good Girls study hard.

۳-۲-۱ تشخیص لحن صدا

تشخیص الگوی آهنگ از توانائی‌های مهم است. با این همه نظر به اینکه دامنه الگوهای آهنگ فوق‌العاده وسیع است و از فردی به فرد دیگر متغیر می‌باشد. درست نیست نظر داوطلب را درباره خصوصیت کلامی شخص جويا شویم. بدین ترتیب، ممتحن ناچار است تا با احتیاط زیاد به آزمون تشخیص آهنگ مبادرت ورزد.

معمولاً دو نوع روش متداول است. شکل اول، ممتحن دو یا چند عبارت را میخواند و از داوطلب میخواهد تا بگوید کدام یک در آن میان تفاوت آهنگی دارد.

Example 8

examinee hears

- ²Why did he ³leave¹ ?
- ²Is he ³there³?
- ²What are you³ doing¹?

examinee reads

1 2 3

Example 9

examinee hears

- ²Where are you ³doing¹ ?
- ³where are you³?
- ²Where are you³ doing¹?

examinee reads

1 2 3

در نوع دوم، داوطلب گفتاری را میشنود و از وی خواسته میشود تا بهترین معنی را از میان سه یا چند گزینه انتخاب کند.

Example 10

examinee hears

Examinee hears

²why is mary³ watching², ²Tom¹?

examinee reads

The speaker is asking

- A) "Why is mary watching tom,"
- B) "Why is tom wotching Mary."
- C) "Tom, why is he watching."
- D) "Tom, why is Mary watching."

۳-۱۰ ادای واج‌ها

آزمون ادای واج‌های تکی و چندتایی برای ممتحن به نسبت تشخیص آهنگ کدام مشکلات بیشتری دارا می‌باشد. اساساً مشکلات بدلیل ماهیت ذاتی پاسخ‌های گفتاری، نتایج ناشی از برگزاری و نمره دادن است. گرچه بهترین روش برای سنجش توانائی فردی در ادای واج‌های زبان بروش مصاحبه در آزمون می‌باشد، این همیشه ساده‌ترین روش نیست. داوطلب، مثلاً، ممکن است ادای صوت خاص یا ترکیبی از اصوات بدلیل آنکه هنوز بر آنها تسلط پیدا نکرده است احتراز کند. بدلیل همین مشکل (و بسیاری معذورات دیگر که بعداً پدید می‌آید) از روش‌های مرسوم دیگری برای وادار ساختن داوطلب به بیان شفاهی استفاده می‌شود. برای اینکار راه‌های مختلفی وجود دارد. روش‌های زیر بخشی از شکل‌های آزمون مناسب و متداول‌ترین آنها است.

۱-۳-۱۰ تقلید

داوطلب گفتار کوتاهی را می‌شنود و آنگاه از وی خواسته تا آنها را بهمان خوبی ادا کند. برحسب هدف آزمون و سطح مهارت داوطلب، حروف صدا دار، دواوائی‌ها، صوت‌های زیر، صامت‌ها، و همگون‌ها، دسته صامت‌ها، استرس و آهنگ را می‌توان سنجید. برای احتراز از مشکلاتی که ممکن است بهنگام تشخیص اشتباهات داوطلبان و نمره دادن با آنها مواجه شویم، بهتر است که هر آیمتی دارای فقط یک نکته آزمونی باشد.

Example 11

- a. blue book
blue book
- b. Visiting people frighten him
Visiting people frightens him
- c. Tom shot the black bird.
Tom shot the black bird.
- d. ²I'm going ³home¹
²I'm ³going home¹.
³I'm ging home¹.

محدودیت این روش این است که توانائی تقلید عیناً از صدائی معین درست بعد از شنیدن آن، با توانائی تقلید همان صدا و بهمان خوبی پس از آنکه دیگر محرک آن صدای معین زائل شده است برابری نمیکند.

۲-۳-۱۰ با صدای بلند خواندن

در این گونه آزمون‌ها داوطلب تعدادی کلمه، جمله یا یک عبارت گفتاری مرتبط را با صدای بلند میخواند و ممتحن تسلط او را بر نکات ویژه‌ای که در متن خواندنی وجود دارد ارزیابی میکند. داوطلب قبل از آنکه متن را با صدای بلند بخواند اول نیم نگاهی به آن میاندازد. مطالب آزمون بایستی مطالب خواندنی برگرفته از زندگی واقعی مثل نامه‌ها، یادداشت‌ها و اعلامیه‌ها باشد. یقیناً از آنجا که این شکل آزمون حاکی از عادت به خواندن هر روزه یعنی - ساکت خواندن - نیست بایستی با توجه به این نوع محدودیت، آنرا بکار برد.

Example 11

examinee hears

|ɔ|
 [θ]
 [1 stress]
 [st]
 [ki]
 [sl]

examinee reads

The children next door often play soccer in the yard and sometimes break my window. Yesterday afternoon, I stayed at home and read a book. After a while, I closed my eyes and went to sleep. A sound at the door made me get up quickly.

۳-۳-۱۰ بازگوئی (واگوئی)

در این شکل از داوطلب خواسته میشود تا داستانی، نکته‌ای، یا قصه‌ای را قبل از امتحان دادن بخواند و سپس آنرا بازگوئی کند. متن خواندنی بایستی بگونه‌ای با دقت گزیده شود که شامل نکاتی باشد که قرار است آزموده شوند. ارزیابی اساساً به آوای واج‌های تک و چند تائی توجه دارد.

۴-۳-۱۰ صحبت کردن درباره تصاویر

عکس‌ها برای تشویق به پاسخ دادن شفاهی است. در این شکل، داوطلب عکس‌ها را ببیند و از وی خواسته میشود تا آنها را توصیف کند یا داستانی درباره آنها بسازد. تعریف داوطلب ممکن است نشان از اشکالات مختلفی داشته باشد؛ در هر حال، ممتحن تنها ب فکر اشکالات تلفظی است.

۴-۱- رهنمودی برای تنظیم سئوالات تلفظی

روش‌های قبل با اشاراتی درباره محدودیت‌های آنها مطرح شد. یک چنین روش‌هایی بنابراین باید با رعایت احتیاط مورد استفاده قرار گیرد و نبایستی بعنوان تنها راه ارزیابی توانایی زبان تلقی گردد. برای اطمینان از سنجشی مطمئن و معتبر از توانایی تلفظی رهنمود زیر بخاطر طرح و تنظیم سؤال و غیره دادن توصیه می‌شود.

۱- انتخاب مناسب مطالب (محرک‌ها) از اهمیتی فوق‌العاده برخوردار است.

این مطالب اول از همه باید شامل زبان محاوره‌ای روزمره با کلمات فراوان باشد. باید نهایت دقت بعمل آید تا داوطلب را از زحمت خواندن و نوشتن زیادی برهاند تا آنها بتوانند تمام حواسشان را به محرک شنیداری جمع کنند.

دوم، تمامی الگوهای استرس (Stress) نمی‌بایستی انتخاب شوند. مؤلف آزمون باید آن نکات را در زبان انگلیسی بشناسد که در زبان مادری داوطلب معادلی ندارد و نمونه‌هایی از آنها را در آزمون بگنجاند. البته، باید خاطرنشان ساخت که چنین سئوالاتی فقط برای گروهی مفید است که زبانشان با زبان آزمون یکی باشد. چنانچه داوطلب زبان مادری متفاوتی داشته باشد، سئوالات می‌بایستی فراگیر باشد، و نمونه‌های سئوالی باید بدون توجه به زبان ویژه‌ای تهیه و تنظیم گردد.

سوم، آزمون واج‌های مجرد از فعالیت‌های زبانی زندگی روزمره بدور است. توصیه می‌شود که نکات تا حد ممکن در متن جا بیافتد. متن زمینه باید مرتبط با کار روزمره باشد. مثلاً، ما بندرت - اگر هم - شعری بخوانیم با صدای بلند خواهدبود. اما اغلب، نامه، یادداشت، خبرها یا اشعار را برای دوستانمان و نزدیکانمان به صدای بلند می‌خوانیم. مطالب درسی از طریق همین منابع میتوانند برای سئوالات آزمون خواندن بصدا بلند، فراهم بشوند.

سرانجام اینکه بایستی در تهیه تصاویر و بعکس‌ها دقت شود تا که بطور مفید بکار آید. عکس‌ها باید ساده و بدون هرگونه ابهام ناشی از تفاوت‌های فرهنگی یا بلحاظ پس زمینه، سن یا وضعیت (خاستگاه اجتماعی - اقتصادی) باشد.

۲- آزمون‌های بیانی بایستی بطور انفرادی در مورد داوطلبان بعمل آید. مگر ممتحن دارای لابراتوار زبان مجهز به ضبط پاسخ‌های گفتاری داوطلبان باشد. در مورد اول ارزیابی پاسخ‌ها الزاماً در همان هنگام اجرای آزمون (و همزمان) انجام میشود، در حالیکه در دومی، اگر پاسخ‌ها خوب ضبط شده باشند به ممتحن اجازه میدهد تا هر زمان که بخواهد ولو با تأخیر به ارزیابی و دادن نمره مبادرت ورزد. چنانچه نمرات همزمان داده شود، معیارهای از پیش تعیین شده برای کار نمره‌گذاری بسیار بااهمیت خواهدبود. ممتحن باید از ارزیابی هر واجی که داوطلب ادا میکند اجتناب ورزد؛ خیلی بهتر است تا ارزیابی به تعدادی نکات ویژه منحصر شود تا داوطلب دقیقاً بداند که به دنبال چه چیزی باید باشد. چنانچه به ترتیب نمره دادن با تأخیر صورت بگیرد، ممتحن فرصت بیشتری برای بررسی واج‌های متعدد

خواهد داشت. و با برگشت به عقب (play back) میتواند تلفظ، فشار ریتم، وقفه (درنگ) همگونی و صداهای زیر دقت کند.

۳- آزمون‌های نمره‌گذاری بازگویی نیازمند معیار یا ملاک‌اند. دقت و صحت آوایی با استفاده از داوران زبان مادری بطور ایدالی میتواند مورد سنجش قرار گیرد. در حالیکه قبول شدن در تست واجی زبان یکمک ممتحنی که به غیر زبان مادری تکلم میکند میتواند به‌خوبی سنجیده شود. بعلاوه، برای بالا بردن ضریب اطمینان بهنگام ارزیابی توانائی بیانی داوطلبان نیاز به حداقل دو ممتحن وجود دارد.

Activities

نمونه‌های زیر بعنوان روشی ممکن برای آزمون تلفظ توصیه میشوند.

A.

examinee hears
Say aloud in English the
Equivalent of /maedrese/

examinee reads
School

B. listen carefully to the following text which will be read only one. When it is finished say it back:

Mr. Smith was feeling so sick that he visited the doctor immediately.

C. Study each item below and then choose the word or the syllable you would stress:

<u>Said</u>	<u>head</u>	<u>net</u>	<u>seat</u>
A	B	C	D
<u>pho</u>	<u>to</u>	<u>gra</u>	<u>phy</u>
A	B	C	D

D. find the vowel sound that is different from the others in the set:

Said	Head	Net	Seat
A	B	C	D

E. Which pair of words "donot" rhyme?

A) said, head

B) bought through

C) bate, bait

D) believe, receive

F. Look at the picture and tell whether the sentences you hear about it are true or false:

Example hears

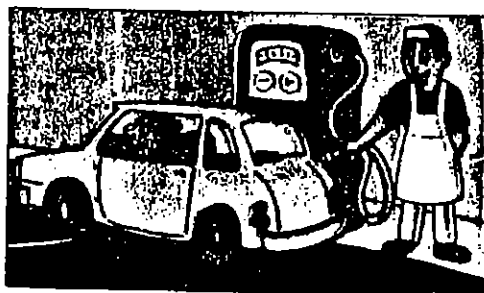
The driver is filling the car.

The man is feeling the car.

examinee reads

T F

T F



2. The following items are designed to test recognition of consonants and vowels of Persian learners of English in the third grade of the Guidance School. Point out the major weakness(es) in each Item. Be specific:
example hears examinee reads

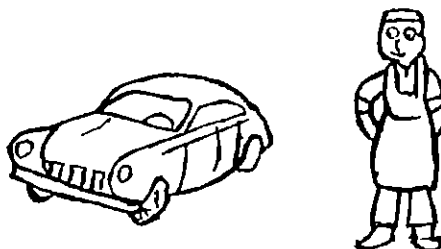
1. Deep	A) Jeep B) Sheep C) Keep D) Deep
---------	---

1. Loiter	A) Lighter B) Loiter C) Later D) Liter
-----------	---

3. Point out the major weakness (es) in the following item. The testee is required to look at the picture and determine whether the sentence he hears is true or false:

example hears examinee reads

I saw him buy the car. T F



4. Find two picture that each illustrates one pronunciation problem. Design one item on each, one testing recognition and the other production. Here is an example.

example hears examinee reads

He's filling the car.	A) He's feeling the car. B) He's filling the car. C) He's feeling the tank. D) He's filling the tank.
-----------------------	--

Or: Where's the driver sitting?

(He's sitting) in the car.

(The examiner is mainly interested in the vowel quality of "car".)



5. Think of other procedures for testing recognition and production of sounds and groups of sounds. What are their advantages/limitations?

آزمون درک مطلب شنیداری

۱-۱۱ مقدمه

درک مطلب شنیداری یکی از عمده‌ترین مهارت‌های زبانی است و نیز دشوارترین آنها. اولاً، گفتگو تنها زمانی میسر می‌گردد، که طرفین قادر به درک هم باشند. گرچه شخص بصور مختلف میتواند کاری کند که دیگران مقصودش را بفهمند (با حرکات دست و صورت). مشکل عمده در ایجاد ارتباط، درک آن چیزهایی است که میشنویم. در حالیکه گوینده بر آن چیزی که میخواهد بگوید احاطه دارد. شنونده بر آنچه که میشنود کنترلی ندارد. بعلاوه، گرچه مؤلفه‌هایی دخیل در امر شنیدن و صحبت کردن یکسانند، شنیدن بر گفتن تقدم دارد و پیش نیاز آن است. تنها پس از آنکه توانش درک آنچه که شنیده میشود میسر گردید فرد فعال میشود و آنرا به سمت جریان ارتباطی پیش میراند. یعنی اینکه، تا وقتی که شخص توانائی درک پیامی را نداشته باشد مطمئناً قادر به ارسال آن نخواهد بود.

با تمام اینها، کار درک کردن، مقایسه، خواندن، نوشتن امری دشوار است. در درک شنیداری، الزاماً فرصت بازگشت یا وقفه وجود ندارد. جریان کلام بدون آنکه منتظر بماند تا شنونده پایه‌پای آن حرکت کند همچنان ادامه دارد. اضافه بر این چون شکل گفتاری بیشتر از نوشتن حاشیه رفتن دارد شنونده محتاج آن است تا بیشتر گوش بزند بماند. زیرا که ممکن است همه آن چیزی را که میشنود نداند. علاوه بر این، مغز (ذهن) آدمیزاد گنجایش معینی برای ذخیره اطلاعات دارد و نمیشود که هم چیزهایی را که میشنود جذب کند. بهمین خاطر، سئوالات آزمون درک شنیداری اغلب شنیدن را با خواندن یا نوشتن میامیزد. داوطلب محرک را میشنود با نوشتن یا علامت زدن گزینه‌های داده شده در دفترچه

آزمون درک خود را نشان میدهد. بدین ترتیب، سؤالات آزمون درک شنیداری پیچیده‌تر از سؤالات آزمون سایر مهارت‌هاست.

منظور از آزمون‌های درک شنیداری این است که توانایی درک و تأویل محرک شفاهی توسط داوطلب سنجیده شود. بویژه، هدف آزمون‌های درک شنیداری برای مبتدیان و متوسطان سنجش درک مؤلفه‌های خاص زبانی نظیر، جملات خبری، سؤالی، و ردوبدل مکالمات کوتاه می‌باشد. آزمون‌های سطوح متوسط به بالا و پیشرفته بهر حال درک گفتار رسمی و غیررسمی و زبان مادری در تمامی شرایط گفتاری را می‌سنجد. داوطلبان در سطوح بالاتر و پیشرفته نوعاً شامل افرادی هستند که علاقمند به فعالیت در محیط‌هایی با برنامه زبان انگلیسی بعنوان وسیله‌ای برای مکالمه هستند.

۲-۱۱ انواع سؤالات

شرایط مطلوب برای ارزیابی درک مطلب شنیداری، از طریق روشی که لازم است که داوطلب به محرک گوش فرادهد و یکی از گزینه‌های شنیداری را که بر مبنای اطلاعات موجود در محرک متکی است برگزیند، میسر است.

Example 1

examinee hears
When do you leave?
A) At six
B) with my wife
C) Yes , I do
D) on main street

examinee reads
A
B
C
D

این سؤال بعنوان «بی‌عیب»ترین سؤال درک مطلب محسوب میشود. زیرا که نیازی به خواندن و نوشتن ندارد. با این وجود این روش اساساً هیچ خوب نیست. بار زیادی را بر حافظه داوطلب تحمیل میکند. مضافاً تجربه نشان میدهد که اگر چنین سؤالی در مورد یک گروه به‌اجرا دربیاید، همین که داوطلبی میکوشد تا گزینه‌ای را که درست است علامت بزند، این حرکت او به داوطلبان کمتر ماهر می‌فهماند که گزینه درست همین الان صورت گرفت. یک نوع آیتم بی‌عیب دیگر شامل عبارات گفتاری‌ای است که داوطلب اجرا میکند. مثلاً از او خواسته میشود تا کارهای زیر را انجام دهد.

Ali, clean the blackboard.

Put your books under the desk

Point to the first chair in the first row.

این‌گونه سؤالات عموماً مناسب کودکان و مبتدیان است. زیرا که شکل آن نیازی به این ندارد که داوطلب بهر شکلی بزبان بفهماند که پیام را دریافته است. با این وجود، محدودیت‌هایی دارد، نخست سنجش درخواست‌ها و اوامر مدنظر است. ضمناً این نوع سؤال خطاب به شخص داوطلب صورت می‌گیرد.

برای خلاص از دست چنین مشکلاتی، آزمون‌های شنیداری معمولی دارای سؤالاتی است که (در میان سایر مهارت‌ها) از مهارت خواندن کمک می‌گیرد. هر سؤالی نوعاً شامل یک محرک کلامی و

گزینه‌های چایی در کتابچه آزمون داوطلب می‌باشد. سه روش ساده و در عین حال مؤثر و عملی برای سنجش مؤلفه‌ها و شیوه زبان و نیز سه شیوه برای سنجش درک کلی در زیر اشاره شده است.

۱-۲-۱۱ عکس‌ها (تصاویر)

تصاویر امکانات فراوانی برای ارزیابی نمونه‌های گفتاری کوتاه ارائه می‌دهند. به داوطلب، مثلاً، عکس نشان داده میشود و از وی خواسته میشود تا بگوید که پاسخ درست یا غلط است.

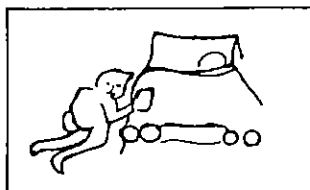
Example 2

examinee hears

The boys washing the car

examinee reads

T F



چنانچه عکس چندین فعالیت را نشان دهد، بیشتر از یک توضیح درباره آن ممکن است. بلحاظ تنوع شکل، ممتحن چهار عکس مرتبط به داوطلب نشان میدهد و از او میخواهد که کدام عکس مطابق با آن چیزی است که میشوند.

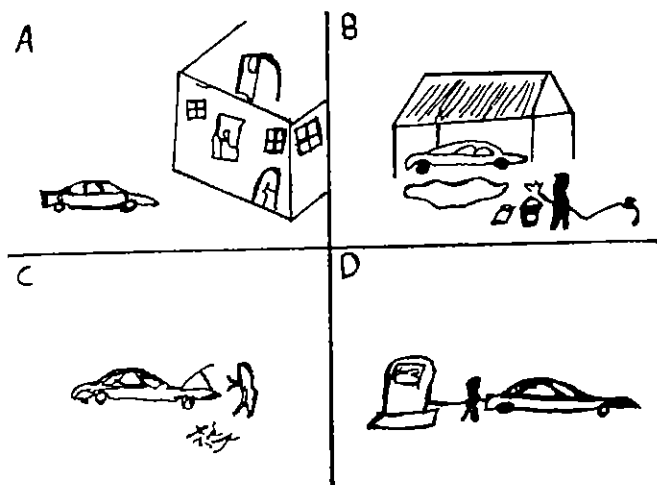
Example 3

examinee hears

The boy's watching the car.

examinee reads

A
B
C
D



عکس‌ها باید آنچنان ساده باشند که باعث ابهام و سوء تفاهم نگردند. البته نیازی نیست که حتماً آثار هنری باشند. یک طرح گرافیکی ساده میتواند وسیله این کار را فراهم سازد.

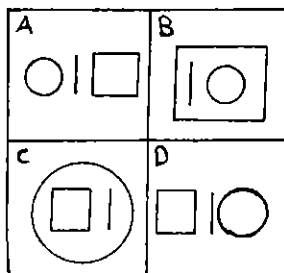
Example 4

examinee hears

A square is to the left of a circle.
there's a line

examinee reads

- A
- B
- C



داوطلبان سئوالی را یکبار می‌شنوند و پاسخی را از میان گزینه سئوالات در دفترچه آزمون برمی‌گزینند.

Example 5

examinee hears

When are you going

examinee reads

- A) Yes , I am
- B) To school
- C) At six
- D) To see Tom

از آنجا که سؤال قرار است درک مطلب شنیداری را بسنجد، زحمت خواندن بایستی به حداقل برسد. بعلاوه تمامی گزینه‌ها بایستی پاسخهایی گرامری و با معنی باشند. به نسبت سئوالهای شفاهی باید ساده‌تر باشند.

در این نوع سؤال، داوطلب عبارتی را می‌شنود (یکبار) بعد در مقابل یک عبارت از چهار گزینه نوشته شده انتخاب میکند.

Example 6

examinee hears

John would've
Passed the exam
If he'd studied
Hard

examinee reads

- A) John passed the exam
- B) John failed the exam
- C) John studied very hard
- D) John would study hard

از میان انواع این نوع سئوالات، داوطلب عبارتی شنیداری را با تأکیدی خاص و بگونه‌ای می‌شنود، بعد از میان گزینه‌هایی که کامل کننده سخن گوینده است یکی را برمی‌گزیند

Example 7

examinee hears
John didn't
Passed the test

examinee reads
A) but he should have
B) but he should
C) but he did
D) but he is.

۴-۲-۱۱ دیکته

دیکته نوشتن شاید روش خوبی برای درک مطلب شنیداری باشد. در یک آزمون دیکته، ممتحن عبارتی را برای داوطلبانی که هر چه را میشوند، مینویسد میخواند. برای اینکه این کار را کامل انجام دهد، داوطلبان نیازمند درک معنایی عبارت، بطور کلی و نیز درک مؤلفه‌های آن و ارتباط بین آنهاست.

۵-۲-۱۱ دیالوگ

شکل گفتاری متنی زبانی به زندگی واقعی بمراتب نزدیک‌تر از سایر روش‌های پیشین می‌باشد. داوطلب مکالمه بین دو نفر را می‌شنود. صدای سومی آنگاه درباره دیالوگ سؤال میکند.

Example 8

	examinee hears	examinee reads
(man's voice)	"Did you see the news On Channel 1 last night?"	A) She was at the movies. B) She was watching Another program.
(woman's voice)	"No, I was so tired that I wanted to sleep, But the movie on Channel 2 was so Interesting that I gave Up sleeping."	C) She had slept very Early. D) She had forgotten to Watch it.
(third voice)	"Why did the woman miss the News?"	

آیتم بالا یک سؤال درک مطلب سرراست است، پرسش‌هایی هم ممکن است مبنی بر توانایی درک و استنباط وی درباره دیالوگ پرسیده شود.

Example 9

	examinee hears	examinee reads
(woman's voice)	"I think I'll Have roast Beef."	A) He can't pay for it. B) He isn't hungry enough. C) He's on a diet. D) He's a vegetarian.
(man's voice)	"Good for you. If I'd your Figure, I too would have Such tasty foods, But I'll just order	

Salad."
(third voice) Why won't the man
Eat roast beef?
Or: "Where are these
people?"

- A) At home.
- B) In a restaurant.
- C) In a doctor's office.
- D) At a grocery.

۶-۲-۱۱ سخنرانی و گفتگو

درک یک سخنرانی و گفتگوی رسمی و غیررسمی یک مهارت حیاتی برای دانش آموزانی است که علاقمند به مطالعه در فضائی هستند که انگلیسی وسیله‌ای است برای آموزش شکل کامل برای کمک به ارزیابی یک چنین درکی شامل سخنرانی دانشگاه هرگونه، یا گفتمان کوتاه با سئوالات چند گزینه‌ای می‌باشد. همانطور که داوطلبان به سخنرانی هفت تا ده دقیقه‌ای بطور طبیعی گوش فرامیدهند، میتوانند از نکات مهم یادداشت بردارند. پس از سخنرانی، بایست به یک سری سؤال درباره موضوع و در رابطه با نکاتی که یادداشت برداشته‌اند، در صورتی که بخواهند پاسخ گویند. آنچه که در زیر آمده نمونه یک آزمون است که برای سنجش توانائی داوطلبانی به دنبال یک گفتگوی کوتاه انجام گرفته است.

"I am now going to tell you about a piece of news reported recently of people who can detect colors with their fingers and even see through solid doors and walls. I will start with the story of an eleven-year-old girl who has normal vision but who can also perceive things with different parts of her skin, and she can also see things through solid walls. Let me tell you how this ability was first noticed. One day she went to her father's office and happened to put her hands on the door of a locked closet. Suddenly she asked her father why he kept so many old newspapers locked away there. Her father did not believe her, so she had to describe the way they were piled up to convince him. Her father brought the girl's talent to the notice of a friend who was a scientist. The scientist gave her a series of tests. Can you believe what these test showed? Well, in one of these tests she was able to read a newspaper through a solid screen. In another test, although she was wearing stocking and slippers, she was able to perceive the outlines and colors of a picture which was hidden under a carpet. Other experiments showed that when the girl's eyes were covered, her knees and shoulders had a similar sensitivity as well. It was also found that although she could perceive things with her fingers, this ability ceased the moment her hands were wet."

examinee hears

1. Who first discovered the girl's abilities?

2. What's the topic of the talk?

examinee reads

- A) Her father.
- B) The scientist.
- C) A reporter.
- D) She herself.
- A) Abilities of young girls.
- B) Talents of girls.
- C) Newspaper reports.
- D) Scientific experiments.

3. The girl has all these abilities except
.....
4. The girl is unable to use her abilities
wet
5. What was the girl able to do in the
experiment with the picture?
- A) seeing through wall
B) perceiveing with fingers
C) reading through a solid screen
D) doing scientific experiments
- A) her skin is wet
B) her eyes are covered
C) she is wearing stockings
D) she has too much clothes on
- A) Name the artist.
B) Recognize the colors.
C) Tell where it was.
D) Say whose it was.

بهنگام انتخاب متون بمنظور آزمون گفتگو - سخنرانی کوتاه، توصیه میشود که متن برای داوطلبان آشنا نباشد، تا با رغبت بدان گوش فرادهند. متن از یک سخنرانی رسمی گرفته تا یک کپ اتفاقی و ملاقاتی رو در رو و تصادفی، پیامی تلفنی، رادیویی و نیز مکالمات تجاری، بازرگانی باید حالت گفتاری معمولی داشته باشد. ضمناً توصیه میشود که آزمون‌های درک مطلب شنیداری شامل سه تا پنج مینی سخنرانی یا گفتار کوتاه که نشانه‌ای از موضوعات و سبک‌های مختلف است، باشد. این روش بر آزمون‌ی با یک تک گفتار بلند مزیت دارد.

در طرح سؤال موضوع آزمون باید معین باشد و گزینه‌ها باید به‌نحوی نوشته شوند که مستلزم توجه و دقت داوطلب به محرک شفاهی باشد.

۷-۱۱ پیام‌های با سروصداهای مزاحم

روشی که بتواند مهارت زبان انگلیسی را در زندگی واقعی بسنجد شامل مطالب شفاهی با افکت‌های صوتی در پس زمینه می‌باشد. داوطلبان جملات مجدد را همراه با صوت اضافه شده به آزمون میشنوند. از آنها خواسته میشود تا آنچه را که میشنوند یادداشت کنند. روش دیگر شامل عبارتی است متشکل از جملات مرتبط، داوطلبان عبارت را میشنوند و سپس به سئوالات درک مطلب چند گزینه‌ای پاسخ میدهند. در عمل این روش از شمار روش‌هایی است که در بالا اشاره شد و به‌همراه صدائی مزاحم که در پس زمینه نواخته میشود، که گاهی شبیه چیزی است مثل صدای امواج رادیو یا سروصدائی که انگار همین حالا (فی‌البراهه) از یک خیابان شلوغ پخش میشود.

این نوع گزینه بر این فرض استوار است که تسلط بر درک شنیداری یکمک توانائی اشراف بر یک سری گفتار مرتبط در حالت صحبت تأیید میگردد. تجربه نشان میدهد که حتی فارغ‌التحصیلان موفق برنامه‌های EFL معمولاً در درک زبان تحت‌شرایط نامتعارف مشکل دارند. در واقع، آنها اعلان‌های خبری در فرودگاه‌ها، یا پیام‌های تلفن را نمیتوانند درک کنند. در صورتی که افرادی که به زبان مادری

خود یعنی انگلیسی تکلم میکنند، با تکیه بر درک خود متن و زیر و بم آن (دقایق آن) سروصداهای پس‌زمینه را جبران می‌سازند.

۳-۱۱ رهنمودی برای تألیف و تنظیم سئوالات درک مطلب شنیداری

برخی رهنمودهای عام برای استفاده در تألیف انواع سئوالات درک مطلب شنیداری توصیه می‌شود:

- ۱- انتخاب عکس‌ها بایستی با دقت صورت گیرد. درک و فهم عکس‌ها بایستی ساده و آسان باشد. یعنی اینکه، دارای معنایی یکسان برای تمامی داوطلبان باشند. پوسترها و عکس مجلات مطالب مناسبی برای مؤلف به حساب می‌آیند و طراحی‌ها با سبک ویژه‌شان همان چیزی است که داوطلب می‌خواهد.

- ۲- انتخاب نمونه‌های محرک‌های کلامی بایستی با رعایت دقت انجام گیرد تا فعالیت‌های گفتاری نوعاً زندگی گونه واقعی را مطمئناً نشان دهند. هیچکدام از آزمون‌های خواندنی نمی‌توانند محرکی مناسب بحساب آیند. داده‌های واقعی نظیر سخنرانی، مکالمات تلفنی، اعلان‌های مختلف، برنامه‌های رادیویی، برای مؤلفان مطالب متناسبی هستند. مؤلف آزمون، بایست اطمینان یابد که موضوع مطالب در نظر داوطلب موضوعی آشنا نباشد. مطالبی درباره آدمها یا سرزمین‌های خیالی بنظر داوطلبان نیز جالب و مناسب خواهد بود.

- ۳- تألیف سئوالاتی درباره گفتگوها و سخنرانی، بایستی بر مبنای ملاحظات سه‌گانه استوار باشد. نخست، سئوالات باید به درک یا فهم تمامی مطالب متکی باشد. کاری با مطالب جزئی نداشته باشد. دوم عبارت راهنما یا سئوال بایستی اصل موضوع را تعریف کند. خاصه اینکه توجه داوطلبان را به نکات آزمون جلب کند مثلاً توجه کنید، سئوال زیر بر گفتار نمونه در این فصل مطرح شده است.

The girl

- A) often visits her father's office
- B) can write outlines with hands
- C) whose story we listened to is 13
- D) has more sensitive knees than finger

در این سوال، صورت مسئله مطرح یا تعریف نمی‌شود. داوطلب بخاطر ابهام سئوال دچار سردرگمی است. او ناچار است تا تمامی گزینه‌ها را قبل از آنکه بداند که چه چیزی از وی خواسته شده بخواند. سوم، گزینه‌ها بایستی بنحوی نوشته شوند که داوطلبان ناچار باشند بدقت به مطلب گوش سپارند. و بالاخره، سئوالاتی که نیازمند به معلومات یا اطلاعاتی خارج از حیطه آزمون گفتاری هست بهتر است از آنها چشم پوشید.

مثلاً صورت مسئله‌ی زیر منشاء مشکل است، زیرا که مطلب گفتاری اسامی گوینده‌ها را مشخص نمی‌کند. بعلاوه، گرچه گزینه‌ها "A" و "C" دارای عدد «۴۰» می‌باشد که ممکن است اشتباهی بجای «۱۴» گرفته شود پس گزینه‌های موجه‌ای برای این متن خاص نیستند و می‌توانند براحتمال حذف گردند.

examinee hears

- (1st voice) "When was your mother born?"
 (2nd voice) "In nineteen forty."
 (1st voice) "On what day?"
 (2nd voice) "On march fourteen."
 (3rd voice) "When was Ali's mother born?"

examinee reads

- A) On March 40, 1914.
 B) On March 14, 1914.
 C) On March 40, 1940.
 D) On March 14, 1940.

۴- گزینه‌ها باید به‌نحوی تألیف و تنظیم گردند که به‌نظر داوطلبان کمتر ماهر که هنوز مطلب را کاملاً نگرفته‌اند، درست (صحیح) برسد. در صورتی که داوطلبان کلمات "Where" و "When" یا "Why" را باهم اشتباه بگیرند یا اگر کلاً علامت سئوالی (wh) را نداند گزینه‌های سئوال زیر صحیح بنظر خواهد رسید.

به‌نگام تهیه گزینه‌ها، توصیه می‌شود تا کوتاه و بلحاظ گرامری یا لغت مشکل نباشد. مضافاً اینکه گزینه‌ها نیز باید منطقی باشند.

۵- مطلب گفتاری هم می‌تواند (توسط ممتحن) بطور زنده خوانده شود یا روی نوار پخش شود. اجرا اولی ساده و نیز در صورت اینکه خاص فقط یک آزمون باشد عملی است.

با این حال دومی ترجیح داده می‌شود، چرا که وقتی در شرایط مختلف، آزمون به‌اجرا درمی‌آید، ضریب اطمینان آن همچنان حفظ می‌شود. در واقع، صدای ضبط شده زمانی که گوینده زبان اصلی برای خواندن متن بحالت زنده در اختیار نیست، ایده‌آل‌ترین است. در هر حالی، آنچه که داوطلبان می‌شنوند، وقتی با سرعتی عادی بیان می‌شود، با اشتباهاتی که گوینده واقعی از همان ابتدا مرتکب می‌شود، مکث‌ها، تکرار و بازگویی و پرت شدن از موضوع، بسیار طبیعی‌تر جلوه می‌کند.

فعالیت‌ها

۱- مزیت یا محدودیت‌های هر نوع سئوالی که در این فصل مورد بحث قرار گرفت چیست؟

۲- آنچه که در زیر می‌آید روش‌های پیشنهادی دیگری است برای آزمون درک شنیداری. نظر شما

درباره هر یک چیست. فواید و محدودیت هر کدام چه هستند؟

A. For instance, in one procedure the testee hears a question or a dialog in English. Then he answers a question in his native language. The following is an example designed for Persian speakers:

examinee hears

How far is it to

Shiraz?

examinee reads

الف) ۸۰۰ کیلومتر.

ب) در جنوب تهران.

ج) نه، دور نیست.

B. The examinee hears a sentence spoken in Persian. He has to give the English equivalent for the sentence:

examinee hears
/dare mahiro gaerm
Mikone/

examinee reads
He's heating the fish.

C. The examiner describes a familiar place, person, object, etc. The examinee listens and then writes down the name of what was described.

D. Testees hear a dialogue in which certain words are missing. Again they hear it a second time with pauses added in place of the missing words. They should write what they think is missing for each pause.

فصل دوازدهم

آزمون بیان شفاهی

۱۲-۱ مقدمه

صحبت کردن بلحاظ آنکه بقصد تبادل افکار با شنونده صورت میگیرد یک پدیده اجتماعی است. البته موفقیت در ایجاد ارتباط بهمان اندازه به گوینده ارتباط دارد که شنونده. این بعهدہ گوینده است که با کلامی سلیس و روان توجه شنونده را جلب کند و آنها را به درک افکار خود وادارد. بدین ترتیب آزمون‌های توانائی کلامی سعی بر آن دارند تا تشخیص دهند که آیا داوطلب توانائی مکالمه مفید و مؤثر را در حالت طبیعی زندگی دارد. برحسب طبیعت این مهارت، آزمون‌های گفتاری می‌باید در مورد هر داوطلب به تنهایی تهیه و تنظیم و به اجرا گذاشته شوند. نظر به همین امر، نیاز به وقت کافی برای اجرا و نمره دادن دارند. بعلاوه، امر ارزیابی در آنها به نسبت مهارت‌های دیگر به خاطر نیاز به تمهیدات برای آزمون‌ی دقیق و ممتحنینی مجرب (کارآزموده)، کار بمراتب دشوارتری است. با این وجود آزمون‌های گفتاری بایستی بدون توجه به تلاش و هزینه‌ی لازم تهیه و تنظیم و به اجرا گذاشته شوند و کار ارزیابی‌شان را به سامان رسانند. چون صحبت کردن مهارتی است که در اکثر برنامه‌های آموزش زبان خود حرف اول را میزند.

۱۲-۲ - سنجش‌های غیرمستقیم

توانائی گفتاری به دو روش قابل سنجش است: غیرمستقیم و مستقیم. اولی از طریق فعالیت‌های شبه واقعی (نیمه واقعی) صورت میپذیرد. در حالیکه دومی طی فعالیت‌هایی انجام میشود که تلاش

شده است تا عیناً باز آفرینی همان حالت و شرایط زندگی واقعی باشد. مباحث زیر انواع روش‌های غیرمستقیم و مستقیم سنجش کلامی است.

توانائی گفتاری بطور غیرمستقیم با استفاده از شماری تکنیک‌های مختلف فرافکنی قابل سنجش می‌باشد. یک چنین روش‌هایی بویژه بدلیل آنکه ممتحن میتواند طرف را وادارد تا آنچه‌ی را که میخواهد بیان کند، مفیدند. داوطلب بهیچ وجه نمیتواند حتی از بیان نکاتی که بر آنها تسلط پیدا نکرده خودداری ورزد. بهر صورت این روش‌های غیرمستقیم مدیریت شده تنها در مورد افرادی که دارای مهارت محدودی هستند میتواند مناسب باشد. مباحث زیر برخی از متداولترین روش‌های غیرمستقیم سنجش مهارت گفتاری را نشان میدهد.

۱-۲-۱۲ استفاده از تصاویر

تصاویر ابزاری مفید و مؤثر در زمینه سنجش توانائی کلامی هستند: آنها نیاز به هرگونه محرک خواندنی یا شنیداری را برطرف می‌سازند. بهر حال تصاویر باید دارای موضوعی آشنا و برای داوطلب نیز مقوله‌ای شناخته شده باشند. تصاویر برحسب نوع آنها و تجارب ممتحن میتوانند به روش‌های مختلف مورد استفاده قرار گیرند:

۱- به داوطلب یک کارت چهار عکسه (عکس‌دار چهارتائی) شماره‌دار داده میشود و از وی خواسته میشود تا برحسب شماره‌ای که ممتحن میخواند راجع به عکس توضیح دهد. توضیح اگر درست باشد ممتحن را قادر می‌سازد تا عکس صحیح را از میان سایر عکسها برگزیند.

ممتحن، برای مثال، از داوطلب میخواهد تا عکس شماره چهار را در دسته‌بندی شرح دهد در حالیکه خودش نمیداند کدام عکس است. سبد عکسها از هم تفکیک شده تا ممتحن نداند که به کدام عکس اشاره داشته است. با توجه به عکس شماره ۴، چنانچه داوطلب با تعریفی مثل ریر آنرا توضیح دهد که ممتحن را قادر سازد تا بدرستی عکس را بشناسد یک امتیاز دریافت خواهد کرد:

“there’s a man wearing a suit and sitting in an armchair. To his right, there’s a table on the floor ..”

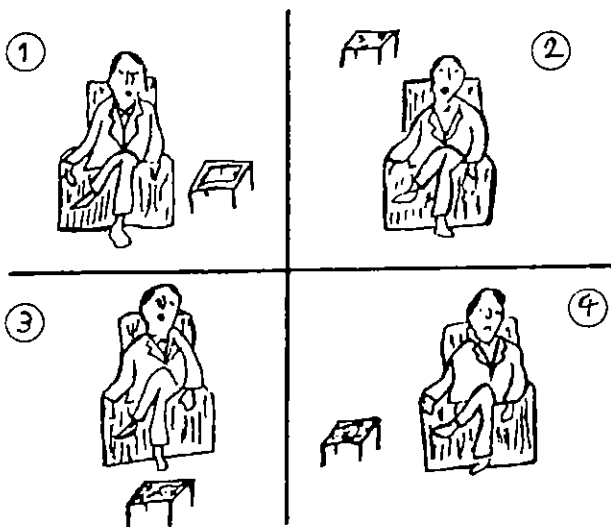
در یک نوع دیگر این روش به فرد یک کارت چهار عکسه داده میشود و آنقدر سؤال میکند تا بفهمد که کدام یک از چهار عکس نزد ممتحن است. سؤالات آری - خیر با یک چنین عکسهائی جور درمی‌آید، اما عکسهای پیچیده‌تر به سؤالات wh گونه نیز نیازمند است:

Example 1

There is a man wearing a suit and sitting in on arm chair.

To his right, there is a table on the floor.....

در یک نوع دیگر این روش به فرد یک کارت چهار عکسه داده می‌شود و آنقدر سؤال می‌کند تا بفهمد که کدام یک از این چهار عکس نزد ممتحن است. سؤالات آری - خیر با یک چنین عکس‌هایی جوردر می‌آید، اما عکس‌های پیچیده‌تر به سؤالات wh گونه نیازمند است.



۲- یک رشته عکس به داوطلب سپرده میشود و از وی میخواهند تا براساس آنها داستانی روایت کند. این نوع سنوآل به این خاطر مفید است که میگذارد تا صحبت بطول بیانجامد. مثلاً، ممتحن عکس زیر را به داوطلب میسپارد و خود داستان زیر را روایت میکند.

Tom and Mary needed some books, so they decided to go shopping together. It was about 2:30 last Monday that they

سپس، متوقف میشود از داوطلب میخواهد تا دنباله داستان را ادامه دهد. ممتحن هر جا که ببیند که روال قصد دارد برای داوطلب عادی میشود، بدان نکته‌ای میافزاید و آنرا داغ میکند. اگر ممتحن احساس کند که داوطلب نکاتی را گذاشته و رد شده به سنوآلاتی میرسد تا آن نکات را واگویی کند. در واقع، اگر داوطلب از سرعنصر زمان براحتی بگذرد، ممتحن با پرسیدن از او به یادش میآورد:

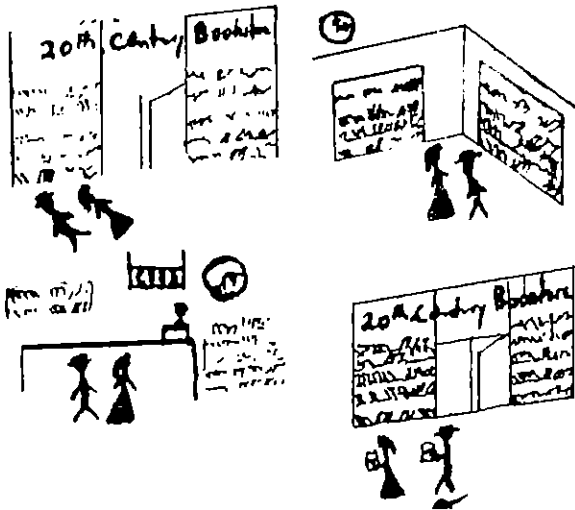
How long did it take them to do the shopping?

Why did they stay at the bookstore so long?

Did they have to wait in line to pay?

دخالت ممتحن اهمیت دارد. چرا که اطلاعاتی که در محرک تصویری وجود ندارد ارائه میدهد و چهارچوبی برای آنچه که میخواهد محک زند آماده میسازد. چنانچه ممتحن علاقمند به استفاده از زمان حال باشد، میتواند داستان را با همان زمان حال شروع کند.

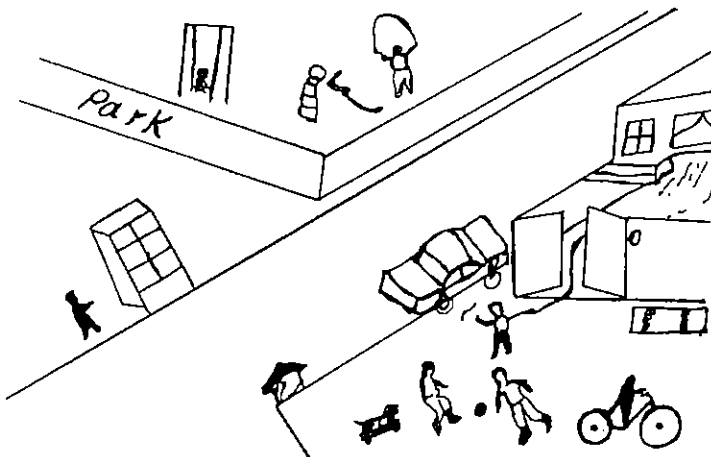
مثال ۲



۳- به داوطلب عکس داده میشود که حاکی از شیئی یا / فعالیتی است. سپس از وی پرسیده میشود که چه چیزی در آن میبیند.

مثال ۳

در اینجا پاسخهای درست اینها باید باشند:



"A man is washing his car."

"Another is walking to the telephon booth."

"Some children are playing in the play ground."

ممتحن میتواند سئوالات مخصوص مطرح سازد. مثلاً، ممکن است بپرسد:

"Where's the man washing his car?"

"What do you think he is going to do next?"

"What do you think will happen if a policeman arrives?"

"What is to the left of the slide?"

۱۲-۲-۲ تغییر شکل دادن

در این گونه سؤال، داوطلب برحسب دستور ویژه‌ای مطالب ارائه شده را تغییر میدهد.

مثال ۴ این جمله را به زمان آینده تغییر دهید

Examinee hears

He works here

today

or



or



Examinee responds

He'll work here (tomorrow).

He will work here (to morrow).

He's going to work here (tomorrow).

۱۲-۲-۳ تبعیت از فرامین

در این گونه سؤال ممتحن از فرامینی که داده میشود باید تبعیت کند.

مثال ۵ این جمله را به زمان آینده تغییر دهید.

Examinee hears

He works here today.

or

or

Examinee executes

Close the door, please.

Where do you live?

Please go ahead or: After you.

۱۲-۲-۴ بازگوئی (واگوئی)

از داوطلب خواسته میشود تا داستان، خاطره، نکته یا نظایر آنرا که بتازگی شنیده یا خوانده است بازگوئی کند. این تکنیک همچنان مفید و مؤثر است. با این حال نظر به اینکه مهارت‌های دیگری نظیر حفظ کردن، مهارت خواندن و انتظام بخشیدن به مطالب را هم میسجد بایستی با احتیاط عمل نمود.

۱۲-۲-۵ شرح دادن

در این روش، از داوطلبان خواسته میشود تا به شرح اشیاء، اعمال و فعالیت‌ها و نظایر آن بپردازد.

مثال ۶

Examinee hears

What's a typewriter?

What's iugging?

Examinee explains

It's a Machine for writting letters on paper using the fingers on a key board.

It's running slowly and steadily.

در صورت وجود داوطلبانی با مهارت بیشتر، ممتحن میتواند سؤالات خیلی مشکلتري بپرسد. سؤالات بایستی به مسائلی مربوط باشد که برای تمامی داوطلبان شناخته شده است. میتوان

پرسشهای ویژه‌ای بسته به علائق افراد و در حد کلماتی که آموخته‌اند بعمل آورد. در اینجا سئوالاتی بطور عام و خاص مطرح شده است.

مثال ۷

How did you register for the courses this semester?

How would you sympathize with an old friend whose mother has just died?

How do you start a car? What about in a cold winter?

How do you make a right turn?

How do you make a call from a public telephone booth?

How do you change a flat tire? What pre cautions do you take at night?

۶-۲-۱۲ صحبت‌های کوتاه

بازهم با حضور داوطلبانی که مهارت بیشتری دارند، ممتحن میتواند با آنها درباره موضوعی بخصوصی برای مدتی معین به گفتگو بنشیند. یا ممکن است راجع به تعطیلات پایان هفته، هر فیلمی که اخیراً دیده‌اند، هر نوع فعالیت جالبی که انجام داده‌اند یا رویدادی که با آن مواجه بوده‌اند به شرح و تفصیل بپردازند.

۷-۲-۱۲ نقش بازی کردن

در صورتی که لازم باشد تا ممتحن یک گروهی از داوطلبان را در یک زمان مورد سنجش قرار دهد، میتوان به ایشان نقشهائی برای بازی در یک نمایش یا در داستانی که ممتحن خود گذاشته است تا واگذار.

۳-۱۲ روش‌های مستقیم

گفتمان یک پدیده اجتماعی است؛ بصرف صحبت کردن حتی بطور دقیق و از روی حقیقت نمیتوان گفت که گفتمانی صورت گرفته است. زیرا روش‌های غیرمستقیم بواقع منعکس کننده موقعیت‌های واقعی زندگی نیستند. مشاهده مردم بهنگام فعالیت‌های روزانه و گوش دادن به صحبت‌های آنها ایدال‌ترین نوع سنجش مهارت زبانی است. بهر حال، یک چنین ارزیابی اگر نه غیرممکن اما خیلی دشوار است. بجای آن میتوان از مصاحبه شفاهی برای آزمون که ابزار واقع بینانه‌ای برای ارزش مهارت گفتاری کلی یک فرد در حالت طبیعی صحبت کردن می‌باشد استفاده کرد. یعنی اینکه، مصاحبه بعنوان آزمون معتبرترین نوع گفتاری است. در این روش یک مصاحبه شونده و ترجیحاً دو یا سه مصاحبه کننده وجود دارند. یکی مصاحبه را هدایت میکند در حالیکه آن دیگری گوش میدهد و نفر سوم یادداشت برمی‌دارد. بعد از سئوالات آشنا و ساده که مصاحبه را گرم میکنند، سئوالات دشوارتر شروع میشود تا جایی که معلوم شود که مصاحبه شونده بدون اشکال دیگر قادر به صحبت کردن نیست

مصاحبه‌گر سعی میکند که مرتب توی صحبت نرود و فقط مشوق مصاحبه‌شوندگان باشد تا بی‌پرده و آزاد صحبت کنند.

مشروح مصاحبه برای ارزیابی‌های بعدی هم بطور ایدال ضبط میشود. بهر صورت از آنجا که ضبط کردن کاری نیست که همیشه ممکن باشد، در پایان مصاحبه ارزیابی صورت میگیرد، در هر حال این امر مستلزم آنست که ممتحنین براساس ملاک‌های از پیش تعیین شده به ارزیابی بپردازند. براساس هدف و معیار تعیین شده نمره دادن هم بروش کلی است و هم بطور جدا جدا. در روش اول یک داوری بطور سرجمع صورت میگیرد که براساس آن مصاحبه شونده، نمراتی مثل «عالی» «خوب» «متوسط» یا «ضعیف»، قبول / رد دریافت میکند. در روش دوم، از طرف دیگر، کار مصاحبه شونده بطور مجدد بر مبنای لهجه، گرامر، واژگان، سلیس و روان بودن و درک کردن سنجیده میشود. این اندازه‌گیری سپس جمع زده میشود تا نمره نهائی بدست آید.

جدول ۱-۱۲ ملاک اندازه‌گیری برپایه سیستم EFL برای استفاده در سنجش مکالمه افراد را نشان میدهد. نمره دادن به مصاحبه‌کنندگان بر شالوده این پنج معیار که مکالمه هم بخشی از آن است، آنگاه بر طبق ملاک‌های جدول ۲-۱۲ سنجیده خواهدشد. بدین‌دلیل که گرامر، واژگان و درک مطلب مهمترین مؤلفه‌های مکالمه شفاهی هستند. لهجه برعکس ترازوی منفی دارد، زیرا که هر چقدر فرد مهارت یافته‌تر باشد بهمان اندازه لهجه کمتری خواهدداشت. شواهد آماری حاصل از تحلیل‌های همبستگی چندگانه معلمان را در این عقیده راسخ‌تر میسازد که به برخی مؤلفه‌ها اهمیت بیشتری داده شود.

برای محاسبه نمره کل مصاحبه شونده. وجود هر یک از پنج مؤلفه یعنی میانگین آن برای دو مصاحبه شونده - طبق جدول اندازه‌گیری بدل به نمره واقعی میشود. مثلاً تصور کنید که مصاحبه شونده‌ای حد میانگین هر یک از پنج مؤلفه عدد سه باشد. این میزان معادل نمره ۵۰ (۵۱) می‌باشد.

توزین	درجه‌بندی	مقیاس
2	3	لهجه
19	3	گرامر
12	3	لغت
6	3	فصاحت
12	3	درک مفاهیم
51		

نمره کل

 رهنمودهایی جهت تعیین درجه

عبارات رفتاری	امتیاز	مقیاس
تلفظ قابل قبول واجی	۶	لهجه
تعداد کمی خطای واجی که هرگز مانع درک نیست	۵	
خطاهای واجی گاه و بیگاه که دقت گوش کردن زیادی را می‌طلبند	۴	
خطاهای واجی مکرر که تکرار را می‌طلبد	۳	
خطاهای واجی پیوسته که درک معانی را مختل میکند	۲	
خطاهای جدی، که درک را عملاً غیرممکن می‌سازد	۱	
تقریباً بدون خطا	۶	ساختار
فقط تعدادی خطاهای جزئی	۵	
خطاهای جزئی گاه و بیگاه که مشکلی در درک ایجاد نمی‌کند	۴	
خطاهای مکرر که گاه و بیگاه در معنی اختلال ایجاد می‌کند	۳	
خطاهای پیوسته که در درک معانی اختلال ایجاد می‌کند	۲	
خطاهای جدی که درک را عملاً غیرممکن می‌سازد	۱	
استفاده مناسب و وسیع از لغات در تمام زمینه‌ها	۶	لغت
استفاده به جا از لغات مناسب برای	۵	
بحث در مورد موضوعات کلی و علایق خاص		
استفاده گاه و بیگاه از لغات نامناسب که	۴	
در هر حال پیام را تحت تأثیر قرار نمی‌دهد		
استفاده مکرر از لغات نامناسب که پیام را تحریف می‌کند	۳	
استفاده دائم و پیوسته از لغات نادرست و استفاده از لغات محدود	۲	
عدم کفایت واژگان اساسی	۱	
صحبت روان و سلیس مانند زبان مادری	۶	فصاحت
صحبت طبیعی و مداوم با مکث‌هایی در نقاط غیرطبیعی	۵	
صحبت روان با مشکلات گاه و بیگاه	۴	
مشکلات مکرر که در فصاحت و روانی اختلال	۳	
ایجاد می‌کنند و کوشش بیشتری را می‌طلبند		
صحبت آهسته، مردد و گاهی اوقات سکوت	۲	
عملاً ناتوان در گفتن جملات متصل به هم	۱	
درک کامل از همه چیز	۶	درک مطلب

- ۵ درک کامل همه چیزی به جز معدودی از محاوره‌ها
یا صحبت‌های تند و سریع یا اصطلاحات غیرمصطلح
- ۴ درک کامل همه چیز اما نیازمند به دوباره‌گویی
- ۲ درک جملاتی که با سرعتی پایین‌تر از حد معمول ادا می‌شود
- ۲ فقط درک صحبت آرام و ساده
- ۱ درک کم حتی از صحبت‌های ساده و آهسته

جدول توزینی ۲-۱۲

درجه‌بندی امتیازات	۱	۲	۳	۴	۵	۶	
لهجه	۴	۳	۲	۲	۱	۰	---
ساختار	۶	۱۲	۱۸	۲۴	۳۰	۲۶	
واژگان	۴	۸	۱۲	۱۶	۲۰	۲۴	
فصاحت	۲	۴	۶	۸	۱۰	۱۲	
درک مفاهیم	۴	۸	۱۲	۱۵	۱۹	۲۳	
						جمع کل	

این روش بسیار عملی است ولی؛ به نوبه خود قابل انتقاد است. بدین ترتیب که:

۱- وقت‌گیر و اجرای آن پرهزینه است.

۲- برای امتحان شونده ایجاد هراس می‌کند.

۳- احتیاج به ممتحنین آموزش‌دیده را دارد.

۴- نمی‌تواند بین امتحان دهنده‌ها از نظر مهارت تمایز قائل شود.

با این وجود به نظر می‌رسد که آزمون مصاحبه شفاهی هنوز هم از نظر صوری معتبرترین و پایاترین و به‌طور معقولانه سنجش عملی توانایی فرد در صحبت کردن می‌باشد. نکته مهم در این است که این سیستم عملی است. به‌علاوه مطالعات تحقیقی پایایی بالایی از بین ارزیابی‌ها یک ارزیابی را نشان می‌دهند و همچنین نشان می‌دهند که هر ۵ مقیاس به‌طور یکسان توزین شده هستند.

۴-۱۲ رهنمودی برای اجرای مصاحبه و ارزیابی آن

مصاحبه شفاهی بمنظور آزمون معتبرترین روش آزمون توانائی گفتاری است. با این همه شکل مصاحبه به خودی خود متضمن معتبر بودن و پایا بودن آن هرگز نیست. هر تست مستلزم تنظیم دقیق، اجرای درست، و نمره‌ای سنجیده است. نکات زیر منابع توصیه و با هدف بهبود مصاحبه‌های شفاهی پیشنهاد می‌شود.

۱- نخست اینکه هر مصاحبه‌ای بایست دارای چهارچوبی دقیق و در حد امکان روانی طبیعی و حقیقی باشد. با وجود این اجرای فی‌البداهه آن مطلوب نخواهد بود. لازم است تا مصاحبه‌کنندگان روی موضوعات کلی قبل از مصاحبه تأمل کنند. آنها بایست قبل از آنکه برای مصاحبه گردهم آیند یک تعداد سئوالات معینی را آماده سازند. توصیه می‌شود تا مصاحبه‌گرهائی که برای نخستین بار به چنین کاری مبادرت می‌ورزند به خود در این باره تمرین بدهند: در این رابطه می‌توانند مصاحبه‌هائی بطور نمونه با افرادی بدل اجرا نمایند یا از مصاحبه‌های قبلی استفاده کنند.

۲- برای دستیابی به نتایج قابل اتکاء بهتر است از حضور دو نفر بهنگام مصاحبه سود برد. نظر دو نفر روی هم درباره ارزیابی کار مصاحبه شونده همیشه از داوری یک نفر تنها بهتر بنظر میرسد. در مدارس، دو معلم همکار در برگزاری کامل یک چنین جلسه‌ای به یکدیگر می‌توانند کمک کنند.

۳- در آغاز مصاحبه با سئوالات ساده می‌توان خاطر طرف مصاحبه شونده را کاملاً آسوده ساخت. وی نباید چنین گمان کند که در مقابل دو قاضی سخت دیگر گرفتار آمده است. از طرف دیگر مدت آزمون باید خیلی کوتاه باشد. مصاحبه‌گر بایست گفتگوئی دوستانه را شروع کند و بتدریج بدل به عضوی شود که با دل و جان گوش می‌سپارد. در حالیکه مصاحبه‌شونده تنها متکلم و جوه می‌باشد. طول هر مصاحبه بستگی به مصاحبه شونده دارد. زمان مصاحبه بین ۱۰ تا ۳۰ دقیقه می‌باشد. اما گاهی ۵ دقیقه هم کفایت می‌کند.

۴- تصمیم مبنی بر استفاده از یک روش خاص یا جهان‌شمول برای نمره‌گذاری بایستی مبتنی بر هدف آزمون باشد. اگر منظور از آزمون شفاهی این باشد که چه کسی قبول و چه کسی مردود شود، آنگاه یک روش فراگیر باعث رضامندی است. از طرف دیگر اگر آزمون مصاحبه بقصد تمایز گذاردن بین افراد مصاحبه شونده باشد، یک روش نمره‌گذاری مخصوص همین کار مطمئناً بکار می‌آید. تجربه نشان می‌دهد که هنگامی که اکثریت مصاحبه شوندگان بلحاظ مهارت در یک سطح قرار دارند. معمولاً استفاده از یک روش فراگیر برای تفاوت گذاری میان آنها چندان معتبر از کار در نمی‌آید. بعلاوه مصاحبه‌گری ممکن است ملاک آسان را بطور کلی بکناری نهند، چرا که اهداف آزمون را شامل نمی‌شود یا ساده‌تر بگوئیم اجماعی وجود ندارد که تلفظ درست ناشی از چه چیزی است. تحت چنین شرایطی یک معیار ویژه و دقیق برای نمره‌گذاری لازم می‌آید. آزمون‌گر بایست اهدافی را که واقعاً و مطمئناً قابل دستیابی است معین سازد.

۵- هر مصاحبه‌ای برای اینکه بعداً بدان نمره داده شود باید ضبط گردد. بدین ترتیب بطور مکرر امکان برگشت برای اینکه هر جنبه‌ای از مصاحبه دایر بر توانائی کلامی مصاحبه شونده مشخص گردد وجود دارد. چنانچه ضبط کردن ممکن نباشد، مصاحبه با طرف بایست پس از ترک جلسه توسط وی مورد ارزیابی قرار گیرد. نمره دادن در حضور وی ممکن است اضطراب برانگیز باشد و بر اجرای کار اثر بگذارد.

عـ امر ارزیابی میتواند ترکیبی در نمرات متغیر بین «عالی» و نظایر آن باشد. یا بر مبنای توضیح مشروح در عمل قرار بگیرد. نمره دادن بروش اول ساده و سریع است؛ دومی بنوعی دشوار است و کند پیش میرود. معهذاً، دومی بیشتر قابل اتکاء است و شمائی از ضعف و قوت نسبی مصاحبه شونده را به دست میدهد.

فعالیت‌ها

۱- در ذیل نمونه‌هایی از روشهای پیشنهاد شده برای آزمون توانایی صحبت کردن آمده است. هر کدام را با دقت مطالعه کنید و استنباط خود را در مورد هر کدام بیان کنید؟ مزایا و محدودیتهای هر کدام چیست؟

الف) به‌عنوان مثال، یکی از روشها مطابق زیر است:

در طی ترم معلم و شاگردانش حدود ۲۰ موضوع یا مطلبی را انتخاب می‌کنند که شاگردان باید نظراتشان را در مورد آن بیان کنند؛ ازدواج؛ فاصله بین نسلها؛ تعالیم مذهبی؛ تصادفهای رانندگی؛ تقلب در امتحانات و غیره. این موضوعات در طی سال در کلاس گاهی به‌صورت مناظره و بعضاً به‌صورت بحثهای گروهی، بررسی می‌شوند. صبح روز امتحان یکی از ممتحنان (دو ممتحن وجود دارد) اسامی شاگردان را می‌نویسد و سپس آنها را در گروههایی به‌صورت اتفاقی طبقه‌بندی می‌کند و لیست این گروهها فوراً ثبت می‌شود. بنابراین هر آزمون شونده فقط نام همگروه خود را در روز امتحان می‌داند. یک دسته کارت با حدود ۲۰ موضوع و همراه آن ۲ یا ۳ سؤال در مورد هر موضوع که قبلاً آماده شده‌اند و همین که گروهها وارد شدند در اختیار آنها گذاشته می‌شوند. آنها دور میز می‌نشینند، کارتی را انتخاب می‌کنند بین خودشان بحث می‌کنند که آیا مایل هستند در مورد این موضوع بحث کنند یا نه، اگر مایل نباشند می‌توانند موضوع دیگری را انتخاب کنند، ولی این انتخاب آخر آنهاست. سپس یکی از اعضاء گروه اولین سؤال نوشته شده در کارت را مطرح می‌کند و بحث شروع می‌شود. ممتحنان هیچ‌گونه دخالتی نمی‌کنند مگر آن‌که احساس کنند که شاگردی تمایل دارد تسلط یابد. یا یکی از دانش‌آموزان حذف شده یا گروه خسته شده و به یک سؤال محرک برای برانگیختن آنها لازم است و یا اینکه باید موضوع دیگری مطرح شود ثابت شده است که با یک گروه چهار نفری تقریباً ۱۵ دقیقه برای این که ممتحنان ارزیابی کلی از هر ۴ آزمون دهنده داشته باشند، کافی است.

ب) آزمون دهنده متنی را که برایش کاملاً ناآشنا است با صدای بلند می‌خواند و ممتحنین عملکرد او را ارزیابی می‌کند.

پ) آزمون شونده ادامه بحثی را که ممتحنین بیان کرده دنبال می‌کند.

آزمون شونده ادامه می‌دهد آزمون شونده می‌شنود

March, April
Tenth, eleventh
Tenth, eleventh

May, June, July, ...
Twelfth, thirteenth, ...

ت) آزمون دهنده به سئوالی درباره تصاویر پاسخ می‌دهد مثلاً با توجه به تصاویر زیر از آزمون شونده پرسیده می‌شود.

“when does the plane leave for Tehran یا when is there a plane for Tehran?”

که انتظار می‌رود جوابهایی نظیر این را بدهد:

“(There’s one) at 9:15”

ث) برای عمل مثل یک مترجم هم‌زمان، دو ممتحن آزمون را اداره می‌کنند. یک ممتحن (وانمود می‌کند که) فقط به زبان مقصد صحبت می‌کند؛ و دیگری به زبان مادری آزمون شونده‌گان صحبت می‌کند. امتحان شونده باید پیام را از یک ممتحن به ممتحن دیگر انتقال دهد، که یکی از آنها به‌عنوان توریست و دیگری به‌عنوان متصدی پذیرش هتل، یکی به عنوان توریست و دیگری کارمند اداره پست، و غیره ظاهر می‌شوند.

ج) از آزمون شونده خواسته می‌شود که خود را در موقعیتی که ممتحن توصیف می‌کند، تصور کند. به او گفته می‌شود تاکنون چه اتفاقی افتاده است. سپس لحظه‌ای مکث می‌شود و آزمون شونده باید آنچه را که احساس می‌کند مناسب است بیان کند. برای روشن شدن موضوع به این مثال توجه کنید:

“You have to return a library book which unfortunately has got torn. What do you say to the librarian when you hand it in?”

خ) اسکارسون (۲۳۴: ۱۹۸۱) مدعی است که: «دانش‌آموزان بزرگسال عموماً قادرند که ایده‌های کاملاً درستی را در مورد عملکرد زبان خارجی ارائه دهند».

۱- توانایی صحبت خود را با اسنفاده از ابزار درجه‌بندی کلی در صفحه مقابل، ارزیابی کنید؟

۲- دو پرسش تبدیلی و دو پرسش تبعیت از فرامین بنویسید؟

۳- تصویری را پیدا کنید که چند فعالیت را نشان دهد. سه پرسش طرح کنید که نیاز به جوابهای کامل داشته باشند؟

۴- نمایشنامه کوتاهی را انتخاب کنید که توسط ۲ یا ۳ هنرپیشه اجرا شود، در حالی که آنها نمایش را اجرا می‌کنند توانایی صحبت کردن آنها را درجه‌بندی کنید؟

۵- از یک هم‌کلاسی بخواهید که با شما مصاحبه کند. از دستورالعملهای مندرج در جدول ۱-۱۲ استفاده کنید. ارزیابی خودتان را با آنچه که دوستان طبق جدول بدست آورده مقایسه کنید؟ نتایج چقدر به هم نزدیک / متفاوت هستند؟ کدام روش را بیشتر می‌پسندید؟ چرا؟

آزمون خواندن و درک مفاهیم

۱-۱۳ مقدمه

بطور کلی خواندن مهمترین مهارت برای فراگیران زبان می باشد به خصوص برای فراگیران انگلیسی به عنوان زبان خارجی. تا به امروز برای مهارت خواندن هیچ نظریه ای که برای تمامی کارشناسان قابل قبول باشد ارائه نشده است. در نتیجه آزمون خواندن شکل های متفاوتی به خود می گیرد و توانایی های مختلفی چون: از پیش خوانی - تندخوانی - و خواندن و درک مفاهیم را ارزیابی می کند.

آزمون های پیش خوانی سعی در ارزیابی توانایی فرد در تشخیص صوری حروف چاپ شده زبان دارد. آزمون های تندخوانی، سرعتی را که آزمون دهنده مطالب را با دقت می خواند اندازه گیری می کنند. آزمون های خواندن و درک مفاهیم سعی در تعیین میزان توانایی شخص در فهمیدن معانی از مطالب مندرج دارد. در این بخش سه راه کار برای آزمون خواندن و درک مفاهیم ارائه خواهد شد.

اولین تکنیک، مربوط به روش اتخاذ شده در آزمون های سنتی خواندن و درک مفاهیم می شود که بر این فرض استوارند که خواندن از چند ریز مهارت تشکیل یافته است. از این رو چنین آزمون هایی معمولاً شامل متن های کوتاهی هستند که هر کدام مجموعه ای از سئوالات چند گزینه ای درک مطلب به دنبال دارند. این پرسش ها سعی در تعیین میزان توانایی فرد در موارد زیر را دارند:

۱- حدس معانی واژگان از محتوای متن

۲- درک ساختار نحوی متن

۳- دریافت ایده های واضح و ضمنی

۴- دریافت منظور اصلی متن

۵- شناخت لحن و روحیه و منظور نویسنده

۶- تعیین راه کارهای ادبی نویسنده و ...

۷- استنباط و نتیجه گیری در مورد محتوای متن

اگر چه سؤالات آزمون که این ریز مهارتها را ارزیابی می کنند در تمام متون خواندن مشاهده می شوند. ولی تعداد دقیق و انواع آنها از آزمونی به آزمون دیگر متفاوت است. آنها عمدتاً متکی به قضاوت نویسنده هستند. مطمئناً تمامی این ریز مهارتها تنها در یک آزمون خواندن و درک مفاهیم مشاهده نمی شوند. از آنجائیکه آنها به طور متقابل وابسته به هم هستند، ممکن است در ترکیبات مختلف در یک سؤال واحد مصادف شوند. برای مثال پرسشی که در مورد ایده اصلی سؤال می کند ممکن است از داوطلب بخواهد نه تنها جمله عنوان را پیدا کند بلکه ایده های واضح و ضمنی بحث شده در متن را بفهمد. خواننده متبحر تمام این توانایی ها را به طور همزمان در تلاش برای خواندن بکار می برد. در ذیل نمونه ای از آزمون برای داوطلبان متوسط «بالای» انگلیسی به عنوان زبان خارجی EFL می آید.

این آزمون متشکل از پاراگرافی از ۱۳۰ کلمه است (برگرفته از الکساندر ۱۹۶۸:۲۰) که به دنبالش ۷ سؤال درک مفاهیم آمده است. البته تعداد سؤالات زیاد است، و برای پاراگراف کوتاهی مانند آن خارج از حد معمول است و مخصوصاً برای اهداف آموزشی طرح شده اند. معمولاً برای متنی متشکل از ۱۰۰ کلمه، حدود سه سؤال کافی به نظر می رسد اگر تعداد سؤالات کمتر از این حد باشد، برای منظور آزمون کافی نخواهد بود و اگر بیشتر از این حد باشد، طراح ناگزیر است که به سراغ نکته های ریز و پیش پا افتاده برود.

مثال ۱: دستورالعمل

این آزمونی است که نشان می دهد شما چگونه می توانید به خوبی متن انگلیسی را دریابید. اول، پاراگراف را به دقت بخوانید و سپس به سؤالات مربوط به آن پاسخ دهید. برای هر سؤال چهار گزینه داده شده است. حرفی را که معرف بهترین جواب در هر گزینه است انتخاب کنید.

PASSAGE

A good education should, among other things, train you to think for yourself. The examination system does anything but "that". What has to be learnt is rigidly laid down by a syllabus, so the student is encouraged to memorize. Examinations do not motivate a student to read widely, but to restrict his readings; they do not enable him to seek more and more knowledge, but induce cramming. They lower the standard of teaching, for deprive the teacher of all freedom. Teachers themselves are often judged by examination results and instead of teaching their subjects, they are reduced to training their students in exam techniques which they despise. The most successful candidates are not always the best educated; they are the best trained in the technique of working under compulsion.

1. The main idea of this paragraph is that ----- .

- a) examination are not achieving the goals they should.
- b) Teachers are not any longer teaching their subjects.
- c) Exams are used to evaluate both teachers and students.
- d) The education system trains you to think for yourself.

2. According to the paragraph, a high grade on an exam is sign of ----- .
 - a) good education
 - b) thorough study
 - c) dependable exam techniques
 - d) training in test – taking
3. Judging teachers on the outcome of exam has ----- .
 - a) made students look down at examination.
 - b) Lowered teaching standards at schools
 - c) Forced students to study test techniques.
 - d) Resulted in a better educational system.
4. We would keep the writer's message if we replaced "that" (line 2) with ----- .
 - a) it should not
 - b) education does not
 - c) train you to think for yourself
 - d) lay down in a syllabus what to learn.
5. We can infer the paragraph that mainly ----- affect the quality of teaching.
 - a) teachers
 - b) students
 - c) candidates
 - d) examinations
6. We can infer from the paragraph that exams should ----- .
 - a) be just based on syllabuses
 - b) train you to think for yourself
 - c) show which teachers are more effective
 - d) make pupils learn the contents of syllabuses.
7. The writer's tone indicates that he ----- .
 - a) feels no respect for education
 - b) tries to find fault with education
 - c) understands the situation and is pleased
 - d) knows the problems but supports the practice

نوع پرسشها، دو الگوی قابل فهم را که به طور معمول در آزمونهای خواندن و درک مفاهیم استفاده شده است نشان می دهد:

شکل سؤال و شکل جمله ناتمام. به علاوه علیرغم این حقیقت که پاراگراف شامل قسمتهایی نظیر واژگانی همچون «induce» «cram» و «deprive» است، سؤالات آزمون استفاده از لغات با دامنه کاربرد وسیع را ممکن می سازد و سؤالات درک مفاهیم نباید به دانش آزمون شونده از این لغاتی کم کاربرد مربوط باشد. خلاصه مطلب این که پرسشها باید به آزمودن نکات زیر در متن معطوف باشند:

- ۱- ایده اصلی: این که امتحانات به اهدافی که باید برسند نمی رسند.
- ۲- تشخیص عقاید ضمنی: این که سیستم امتحانی منجر به تأکید بر آموزش در جهت آزمون دارد؛ و این که قضاوت معلمین در نتیجه امتحانات، تأثیر نامطلوب بر نحوه آموزش آنان دارد.
- ۳- درک ساختار متن: که امتحانات هر چیزی که نباید انجام دهند، انجام می دهند.

۴- استنتاج و استنباط: این که امتحانات موجب کیفیت ضعیف آموزش در زمان حال هستند؛ و این که امتحانات باید به افراد تعلیم دهند که در مورد خود بیشتر تفکر کنند.

۵- تعیین لحن نویسنده: این که او در مورد نحوه سیستم امتحانی خرده گیر است.

با این که کارشناسان مشخص نکرده اند که چه تعداد پرسش های ریز مهارت های خواندن در این امر دخیل اند و اگر سنجش این ریز مهارتها ممکن باشد، انواع سنتی آزمونهای خواندن هنوز هم به عنوان رایج ترین ابزار سنجش توانایی خواندن می باشد. در حقیقت، این روش چهارچوب مفیدی برای طرح آزمونهای خواندن فراهم می کند. بعلاوه ممکن است تا حدی ارزش دستورالعملهای خواندن را نیز داشته باشد.

دومین شیوه برای امتحان، خواندن و درک مفاهیم روش جای خالی است، که در این روش، با استفاده از متنی منشور با تعدادی کلمات حذف شده و از دانش آموز خواسته می شود تا جاهای خالی را پر کند. به منظور انتخاب جواب درست از میان گزینه های قابل قبول برای هر جای خالی، خواننده باید زبان نوشتار را با استفاده از کلیدهای نحوی، معنایی و عملی بررسی کند. شیوه طرح متون جای خالی در فصول بعدی شرح داده خواهد شد.

در روش سوم برای سنجش خواندن و درک مفاهیم تکنیک استفاده از متون کوتاه است. اساساً این روش، شیوه ای است برای آزمودن توانایی خواندن از طریق استفاده از مطالب خیلی خلاصه که به دنبال آن یک یا دو پرسش درک مطلب برای آزمودن درک کلی معنا می آید. برخلاف آزمونهای سنتی خواندن که خواندن را به عنوان مجموعه ای از چند مهارت می پندارند، آزمونهای متون کوتاه به این نظریه معتقد هستند که «خواندن، یک توانایی واحد و یگانه می باشد». مطالب محرک این گونه آزمونها متشکل از متون معتبری است که نمونه ای از کارهای خواندن آزمون شونده را تشکیل می دهند، مانند کتابهای دانشگاهی، مجلات و تبلیغات دانشگاهی، مقالات روزنامه ها، بروشورها و متن های دانشگاهی، عناوین پزشکی و سرفصل دروس.

به نظر می رسد، این مطالب بیشتر به امور زندگی روزمره نزدیکتر می باشند.

مثال ۲

246 Evaluation of language ability (2)

Either semester. Two class periods. Prerequisites; Measurement 219 which may be taken concurrently. Review of present – day theory. Intended to, increase senior's skill in preparing and assessing classroom tests.

FARHADY, JAFARPUR

According to this description, the students who want to take this course -----

- should take course 219 in the following semester (s)
- ought to be skillful in preparing classroom tests
- must ask professors farhady and jafarpur's permission
- have to be in their third year of studies in college

چنانکه گفته شد تکنیک متون کوتاه به آزمون ساز این اجازه را می‌دهد که بر آزمون درک کلی از مطالب محرک تمرکز داشته باشد تا این‌که فقط به‌طور نامعقولی به تحلیل جزء به جزء متن بپردازد. در این جا هیچ تأکیدی بر پرسشهای استنباطی نیست. سعی آزمون بیشتر در جهت سنجش مهارت خواندن است تا سنجش هوش، استنباط، یا توانایی تفکر، بعلاوه از آنجا که شمار زیادی از مطالب محرک می‌توانند در آزمونهای خواندن گنجانده شوند، تنوع متن‌ها می‌توانند بیشتر باشد. بنابراین مزیتی را که برخی از بافتهای خاص خوشایند برخی از داوطلبین باشد و نه همه آنها، کاهش می‌دهد.

۲-۱۳ رهنمودهایی برای ساختن پرسشهای خواندن و درک مفاهیم

رهنمودهای کلی زیر برای آماده کردن آزمونهای خواندن و درک مفاهیم پیشنهاد می‌شود:

- ۱- در تهیه آزمونهای مختلف خواندن، توصیه می‌شود که برخی نکات را به‌خاطر بسپارید. اولاً، متن باید از لحاظ کاربرد زبانی واقعی باشد. در آزمونهای پیشرفتی، نمونه برداری باید بازتاب موضوعات اهداف یادگیری زبان باشد؛ از طرف دیگر، در آزمونهای مهارت نمونه برداری باید نوعی از فعالیت خواندن را منعکس کند که آزمون دهنده ممکن است بعدها در زندگی روزمره با آن مواجه شود. اگر متن انتخاب شده حاوی برخی لغات ناآشنا باشد، دانستن این لغات نباید برای درک متن ضروری باشد و گرنه متن باید ویرایش شود.

برای کسب اطمینان از تطابق بین متن انتخاب شده با سطح آزمون دهندگان می‌توان از فرمولهای خوانایی استفاده کرد.

در این خصوص فرمولهایی وجود دارد، که توسط متخصصان خواندن برای تعیین سطح خوانایی متن طرح شده است:

برای مثال می‌توان از فرمول فلش^۱، فرمول فرای^۲، فرمول داله^۳ چال و فرمول اسپاچ^۴ نام برد. منحنی خوانایی فرای در ضمیمه کتاب آمده است. استفاده از نتایج این فرمول مانند سایر فرمولها آسان است و نتایج مهمی را دربردارد.

بعلاوه از آنجائیکه محتوا و سازمان بیانی مطالب محرک تأثیر به خصوصی در درک مطلب دارد، آزمونهای خواندن باید شامل متونی باشد که دارای محتوای مناسب و از نظر فرهنگی قابل قبول باشد و معرف دانش و علاقه آزمون شوندگان باشد.

^۱. flesh

^۲. Fry

^۳. Dale - chall

^۴. spache

بهر حال باید از مطالبی که حاوی موضوعات معمولی و آشنا برای آزمون دهنده باشد اجتناب ورزید. علاوه بر این، هر متن خواندن باید به حد کافی دارای محتوا و مفهوم باشد تا طرح تعدادی از سؤالات منطقی و مناسب ممکن باشد. اگر چه تعداد متون و طول هر کدام از آزمونی به آزمون دیگر فرق می‌کند، متون حاوی حدود ۱۰۰ تا ۳۰۰ کلمه مناسب به نظر می‌رسد و در حقیقت طول متن باید متناسب با سطح دشواری و تعداد گزینشهای موجود در آزمون باشد. برای این که آزمون خواندن و درک مفاهیم مؤثر باشد، هر یک از گزینشهای آن بایستی بافت کافی برای پنج پرسش درک مفاهیم داشته باشد.

پرسشهای آزمونهای سنتی یا متن کوتاه می‌توانند به دو گونه باشند: تشریحی و چند گزینه‌ای. طرح پرسشهای تشریحی آسان است، اما نمره‌دهی آنها مشکل است. مسلماً این کار با مجاب کردن آزمون دهنده برای نوشتن پاسخ جهت اثبات درک خود از متن خواندن انجام می‌گیرد. از طرف دیگر، طرح پرسشهای چند گزینه‌ای سخت است، اما نمره‌دهی آنها آسان است. برای طرح پرسشهای چند گزینه‌ای ملاحظات زیر را در نظر داشته باشید: اولاً، تمام گزینه‌ها باید از نظر نحوی درست و از نظر معنایی قابل قبول باشند. بعلاوه آنها باید از نظر طول و اندازه برابر باشند. دوماً، طراح سؤال باید بطور واضح مسئله را روشن کند سؤالی مانند نمونه زیر منحرف‌کننده است، زیرا تا آزمون دهنده تمام گزینه‌ها را نخواند قادر به جوابگویی نخواهد بود.

Teachers ----- .

- encourage students to learn up hastily for exams.
- Make students look down examinations
- Force students to study test techniques
- Help students to learn the art of test - taking

سوماً، سؤال باید دانش‌آموز را مجاب کند که قبل از پاسخ دادن به سؤال تمام مطالب متن را به دقت بخواند. از طرح سؤالاتی که بدون خواندن متن قابل پاسخگویی باشند باید اجتناب کرد. برای مثال سؤال زیر وابسته به متن نیست و حتی آزمون دهنده مبتدی هم می‌تواند پاسخ‌های A, B, D را بدون این که متن را بخواند به آسانی حذف کند.

One can infer from the paragraph that examinations should ----- .

- test what the students have memorized hastily.
- Be important to both students and teachers.
- Motivate students to study extensively.
- Contain only difficult questions.

چهارماً، سؤالات باید آزمون شونده را مجبور کنند تا درک و تعبیر درستی از مطالب خوانده شده داشته باشد. از طرح پرسشهایی که صرفاً بخشهای متن را به هم وصل می‌کنند باید اجتناب کرد.

Exams lower the standard of teaching because ----- .

- the most successful candidates are not always the best educated.
- The best trained in the technique of working under compulsion are students.
- They are reduced to training their students in exam techniques.
- They deprive the teacher of all freedom.

آزمون شونده‌ای که متن را درک نکرده است، قادر است که به این سؤال پاسخ صحیح دهد، زیرا مرتبط کردن لغات سؤال با لغات موجود در متن آسان است.

پنجم، سؤالات نباید هیچ سرنخی در مورد جواب صحیح بدهند. سؤال زیر با سه گزینه نادرست موجب هدایت آزمون دهنده به گزینه درست می‌شود. حال اگر این سؤال در آزمون نمونه (ارائه شده در بالا) گنجانده شود، باعث لو رفتن گزینه درست از بین گزینه‌های دیگر می‌شد.

According to the paragraph, examinations do anything "except" ----- .

- a) training to think for themselves
- b) limit pupils to memorize their tasks
- c) make students study facts hastily
- d) lower the standards of teaching

و بالاخره، از طرح گزینه‌های معمول، مانند «همه گزینه‌های بالا» یا «هیچکدام از گزینه‌های بالا»، «نه C نه D» باید اجتناب شود. این عبارات اغلب چیزی به سؤالات نمی‌افزایند و به راحتی می‌توانند مشخص و حذف شوند.

فعالیت‌ها:

۱- در زیر روش‌های دیگری از آزمون خواندن و درک مفاهیم پیشنهاد شده است. ارزیابی شما در مورد هر کدام چیست؟ مزایا و محدودیتهای هر روش کدام است؟

الف) از آزمون شوندگان خواسته می‌شود که متن را تا حد امکان به سرعت بخوانند. هنگامیکه یکی از آنها متن را تمام می‌کند دستش را بلند کند و ممتحن زمان را یادداشت می‌کند. سپس او تعدادی از سؤالات چند گزینه‌ای را می‌خواند. امتحان شوندگان نمی‌توانند دوباره به متن خوانده شده نگاه کنند.

ب) هر آزمون شونده یک متن ناآشنا را با صدای بلند می‌خواند و سپس به سؤالاتی که ممتحن مطرح می‌کند پاسخ می‌دهد.

پ) به آزمون شوندگان متنی داده می‌شود که در آن چندین کلمه، عبارت، یا جملات نامربوط وارد شده است. آنها بایستی متن را بخوانند و موارد نامربوط را خط بزنند.

۲- متن زیر و پرسشهای آن را بخوانید. ضعفهای اصلی هر پرسش را مشخص کنید. دقت کنید:

The temperature on the surface of the sun is about 11,000 degrees fahrenheit. You would think that the closer you got to the sun, the hotter you would be. Let, just a few thousand feet above the earth, the air is very cold.

The heat from the sun is radiant energy. This energy is transmitted through transparant things such as air or glass. But once radiant energy hits something that does not let light through, it is absorbed as heat. When energy from the sun hits the ground, the ground is warmed. The ground then warms the air just a bove it.

Cities are often warmer than rural areas. Large buildings and cement streets give off more of the sun's energy than does soil and this warms the air. This explains why it is warm near the earth where we live, and why it is cold a few thousand feet above us.

1. From the four choices below, ----- allow(s) light to pass through.
 - a) windows
 - b) bricks
 - c) buildings
 - d) the ground
2. Radinant energy passes through
 - a) buildings
 - b) cement streets
 - c) the air
 - d) the sun
3. Sunlight that cannot pass through an object is ----- .
 - a) transmitted
 - b) sent back to the sun
 - c) absorbed by the object
 - d) changed slowly
4. "It" (line 6) refers to ----- .
 - a) light
 - b) radinat energy
 - c) something transparant
 - d) air
5. The air near the earth is warmed when ----- .
 - a) sun light passes through the air
 - b) an object gives off the heat from the sun
 - c) light is reflected by windows in large buildings
 - d) clouds do not let the sun's heat escape from the air
6. Cities are often warmer than rural areas because ----- .
 - a) rural areas absorb more of the sun's energy
 - b) cement streets do not absorb any of the sun's energy
 - c) large buildings give off a greater amount of the sun's energy
 - d) solid gives off more often sun's energy
7. We can conclude from the passage that ----- .
 - a) the temperture is about 11,000 degrees fahrenheit a few thousand feet above the earth.
 - b) Radiant energy is absorbed by glass
 - c) The closer you get to the sun, the hotter you are
 - d) Rural areas are cooler than cities because the farmers have fewer objects to absorb radiant energy
8. You can guess from the context that "yet" (line 3) means ----- .
 - a) at the time that
 - b) at one time
 - c) just one time
 - d) at a time in the past
9. You can guess from the context that "once" (line 5) means ----- .
 - a) at the time that
 - b) at one time
 - c) just one time

d) at a time in the past

10. You can guess from the context that "rural" (line 9) means the "opposite" of ----- .

- a) urban
- b) agricultural
- c) regional
- d) native

11. "The temperature on the surface of the sun is about 11,000 degrees fahrenheit". In which of the following sentences does about have the "same" meaning?

- a) This book is about the life of king henry VIII.
- b) I've thought about Ali a lot, but I have time to write to him.
- c) Can you meet me at about three o'clock tomorrow afternoon in my office?
- d) I don't know what you are talking about.

12. "Let, just a few thousand feet about the earth ----- ". In which of the following sentences does just have the "same" meaning?

- a) you can't expect her to be too responsible. She's just a child.
- b) My parents are just as happy as I am about my decision to attend the university.
- c) There is a beautiful lake just west of here where we can have a picnic.
- d) Because his decisions are always just ones, the group elected him president.

۳- سه سؤال چند گزینه‌ای خوب راجع به متن که در سؤال ۲ آمده است بنویسید، از کجا می‌دانید که سؤالات شما مربوط به متن هستند.

۴- آیا آزمون شنیداری متفاوت از آزمون خواندن و درک مفاهیم می‌باشد؟ چرا؟

آزمون‌های مربوط به توانایی نوشتن

۱-۱۴ مقدمه

نوشتن مهارت پیچیده‌ای است که موفقیت در آن به چندین روش میسر است: آمادگی برای نوشتن و شروع به نوشتار منجر به نوشتن آزاد می‌شود. به‌طور طبیعی، هر کدام از این دوره‌ها نوع مخصوصی از فعالیتها را می‌طلبد. برای مثال، آنهایی که مخصوص دوره آمادگی برای نوشتن هستند برای تقویت انواع واژگان و الگوهای نحوی خاص، طرح شده‌اند. شروع به نوشتن مربوط به فعالیتهایی است که در برقراری ارتباط برای پیامهای کوتاه تحت شرایط کنترل شده مناسبند. نوشتن آزاد، که آخرین دوره نوشتن است، به فعالیتهایی مربوط است که فراگیر زبان شروع به برقراری ارتباط آزاد در مورد افکار و عقایدش در خصوص موضوعی معین می‌کند.

به‌همین ترتیب، هر دوره نیازمند انواع خاصی از آزمون برای ارزیابی دقت شخصی در استفاده از زبان نوشتاری می‌باشد. در این بخش بر روی تکنیکهایی که برای سنجش آمادگی برای نوشتن، شروع به نوشتار و نوشتن آزاد به کار می‌روند تمرکز می‌کنیم.

۲-۱۴ آمادگی برای نوشتن

در زیر تکنیکهای متعدد آزمون که مناسب برای یادگیری فراگیران انگلیسی بعنوان زبان خارجی است ذکر شده است، این فراگیران توانایی محدودی در یادگیری زبان دارند. این افراد خط را یاد گرفته‌اند، اما هنوز قادر نیستند تا لغات را در جملات قابل قبول نحوی بکار ببرند.

۱-۲-۱۴ رونویسی

در آزمون رونویسی، آزمون‌دهندگان باید مدل نوشته شده را تا حد ممکن باز آفرینی کنند. عملکرد آزمون‌دهندگان بر مبنای تکمیل بودن رونویسی ارزیابی می‌شود. این روش به‌ویژه در این مرحله خیلی مفید است، چون باعث می‌شود افراد دقت بیشتری به مکانیسم نوشتن و کلام متن مبذول دارند. مطمئناً این روش برای کسانی که زبان مادری‌شان سیستم نوشتاری متفاوتی را از زبان مورد آزمون استفاده می‌کنند مفید خواهد بود.

۲-۲-۱۴ هجی کردن

متداولترین تمرین برای آزمون توانایی هجی کردن افراد این است که از آنها بخواهیم کلمات، جملات و یا ترجیحاً متونی را که شفاهاً دیکته می‌شود بنویسند، درست و صحیح هجی کردن، ممکن است برای نوشتن یک فاکتور ضروری نباشد، اما برای اینکه نوشته تاثیر بیشتری بگذارد هجی کردن کاملاً ضروری می‌باشد.

۳-۲-۱۴ نقطه گذاری

به افراد جمله یا مجموعه‌ای از جملات بدون نقطه‌گذاری داده می‌شود، آنها باید تکالیف خود را نوشته و تمامی علامتهای نقطه‌گذاری و حروف بزرگ را بنویسند.

مثال ۱

آزمون دهنده می‌خواند

آزمون دهنده می‌نویسد

Tom had a bad cold consequently

He stayed in bed for two days.

توجه داشته باشید که در اینجا، بسته به سبکی که بکار رفته است بیش از یک پاسخ صحیح وجود دارد. مسلماً اینگونه سؤال زمان قابل توجهی را برای نمره‌دهی می‌طلبد مخصوصاً هنگامی که تعداد آزمون دهنده‌ها نیز زیاد باشند. بنابراین، گونه‌های متفاوت زیر برای افراد زیاد، ممکن است سودمند باشند. در مرحله نخست، از آزمون‌دهنده خواسته می‌شود که علامتهای نقطه گذاری را در درون جاهای خالی قرار دهد. سپس از آنها خواسته می‌شود که جمله علامت‌گذاری شده صحیح را از بین چهار گزینه داده شده انتخاب کنند.

مثال ۲

آزمون دهنده می‌خواند

آزمون دهنده می‌نویسد

Tom had a bad cold

Consequently He stayed

In bed for two days

مثال ۳

a) Tom had a bad cold; consequently he stayed in bed for two days.

- b) Tom had a bad cold consequently he stayed in bed for two days.
 c) Tom had a bad cold, consequently he stayed in bed for two days.
 e) Tom had a bad cold, consequently he stayed in bed for two days.

۴-۲-۱۴ ترکیب جملات

توانایی آزمون‌دهنده در ترکیب جملات، از طریق انواع ساده سوالات ترکیب جملات، سنجیده می‌شود. به مثالهای زیر توجه کنید:

مثال ۴

آزمون دهنده می‌نویسد

آزمون دهنده می‌خواند

Combine the following sentences using that: John bought a newcar last week. He is happy with his car.

یا

I am confident. The teacher discuss chapter Eight the next class session.

۴-۲-۱۴ تقلیل جملات

جمله‌ای به آزمون دهنده داده می‌شود و از او خواسته می‌شود تا قسمت مشخص شده را خلاصه کند. در اغلب مواقع، کلمه کلیدی جهت نشان دادن چگونگی انجام کار، داده می‌شود.

مثال ۵

آزمون دهنده می‌نویسد

آزمون دهنده می‌خواند

David went home and told
His father the news.

Having, David told
His father the news

۴-۲-۱۴ تکمیل جملات

در این گونه سوالات لغات ویژه‌ای، که معمولاً کلمات نقشی می‌باشند، حذف می‌شوند، و از آزمون دهنده خواسته می‌شود که لغت جا افتاده را براساس کلیدهای داده شده در متن پیدا کند. این گونه سؤال می‌تواند در چهارچوب سؤال چند گزینه‌ای نیز مورد استفاده قرار گیرد.

آزمون دهنده می‌خواند

آزمون دهنده انتخاب می‌کند

..... Chapter Eight for the next class session, the professor left the classroom.

- a) While assigning
 b) Having assigned
 c) Upon assigning
 d) As assigned

a, b, c, d

با اینکه نوع سؤال شبیه آزمونهای ساختاری است، اما به یک دلیل اساسی از آنها متفاوت می‌باشد. سوالات نوع ساختاری اغلب مربوط به حالت گفتاری هستند و گزینه‌های انحرافی در انگلیسی

محاوره‌ای قابل قبول نمی‌باشند. با این وجود، نوشتن پرسشها به حالت نوشتاری دلالت می‌کند و گزینه‌های انحرافی، اگر چه احتمالاً در گفتار قابل قبول می‌باشند، در نوشتار مناسب نیستند.

۷-۲-۱۴ تغییر جملات

از آزمون‌دهنده خواسته می‌شود تا جملات را براساس دستورالعمل خاصی تغییر دهند:

مثال ۷

Rewrite the following sentence replacing "he" for "I": I saw the boys going to my house.

Or: Change to the plural: The librarian is putting the book on the shelf:

۸-۲-۱۴ بسط جملات

در این نوع سؤال، پاراگرافی که دارای چندین جای خالی می‌باشد، در اختیار امتحان شوندگان قرار می‌گیرد و از آنها خواسته می‌شود که جاهای خالی را با توضیح مناسبی پر کنند:

مثال ۸

پاسخ

متن

A ... student ... left the class ... as the bell rang. Another student was ... waiting for him outside the ... classroom. As he neared, the other smiled ... and said "Hello"

A "young/tall" student "from" the last row/from behind me" left the class "hastily/just" as the bell rang. Another student was "impatently/anxiously" waiting for him outside the "next/opposite" classroom. As he neared the other smiled "in a friendly manner/sarcastically" and said "Hello"

۹-۲-۱۴ تشخیص خطاها

این نوع سؤال شامل جمله‌ای با برخی لغات و عباراتی است که زیرشان خط کشیده شده است. امتحان شوندگان باید قسمتی که نیاز به اصلاح دارد جدا کنند.

مثال ۹

جواب

متن

^AOn the way to the press
The conference^B Defence
Minister stopped to chat^C
^DWith the president

A, B, C, D

اگر چه این نوع سؤال در نوشتن آزمونها بسیار متداول است، با این حال باید در کاربرد آنها دقت لازم به عمل آید. از نظر آموزشی، کار درستی نیست که فرم زبانی غلطی در اختیار امتحان شوندگان قرار داده شود، بلکه آنها باید مدام با شکل صحیح زبان در ارتباط باشند. بعلاوه، اگر چه این نوع پرسش

اطلاعاتی در مورد توانایی ویرایش امتحان شوندگان در اختیار ممتحن قرار می‌دهد. اما در مورد توانایی نوشتاری آنها هیچ اطلاعاتی را روشن نمی‌سازد.

۱۴-۳ نوشتن کنترل شده

در مورد افراد ماهر، باید پا را از ارزیابی توانایی تشخیص یا تکمیل جملات دستوری فراتر نهاده و توانایی آنها را در انتقال تفکر و اندیشه در نوشتار را ارزیابی کرد. با این وجود، از آنجائیکه امتحان شوندگان هنوز مهارت لازم را در پیچیدگی‌های زبانی ندارند، تمیهدات ویژه‌ای برای کمک به آنها به کار گرفته می‌شود، تا در نوشتار آزاد، اشتباهات کمتری داشته باشند، به ویژه تکلیف کنترل شده شامل، نمونه نوشته شده با دستورالعملهایی برای تبدیل یا تغییرات ویژه زبانی در دوباره نویسی نمونه نوشته شده می‌باشد. بخش بعدی، روشهای متداول این آزمون را نشان می‌دهد.

۱۴-۳-۱ تغییر شکل جملات

در این نوع آزمون، مجموعه‌ای از جملات در اختیار امتحان شوندگان قرار می‌گیرد و از آنها خواسته می‌شود با تغییر زمان، فاعل، حالت، خبر و سایر علائم، این جملات را به یک پاراگراف تبدیل کنند.

مثال ۱۰

آزمون دهنده می‌خواند

آزمون دهنده می‌نویسد

Rewrite the following paragraph changing into active voice:

When the cars stopped, Mary was attracted by the blue car.

۱۴-۳-۲ جمله‌سازی به وسیله کلمات

گروهی از کلمات در اختیار امتحان شوندگان قرار می‌گیرد تا آنها را در یک جمله معنی‌دار به کار برند.

مثال ۱۱

آزمون دهنده می‌خواند

آزمون دهنده می‌نویسد

/ boy- go- home- car- yesterday /on- way- A/ the boy went hom by car yesterday.
his- car- break- ... On the way, his car broke ...

۱۴-۳-۳ جمله سازی از طریق پاسخ و پرسش

مجموعه‌ای از سؤالات به امتحان شوندگان داده می‌شود تا به آنها پاسخ دهند این سؤالات به فرد یا مکانی اشاره دارند که امتحان شونده آنها را به خوبی می‌شناسد یا این‌که، سؤالات براساس نشانه‌های تصویری می‌باشند. پاسخ‌ها به صورت یک پاراگراف درآورده می‌شوند.

مثال ۱۲

Is his name, Ali?
Is Ali 30 years old?
Is he studying English now?
Or
What is his name?
How old is he?
What is he studying?
Or
What is doing?
Is he drinking tea or coffee?
Is he reading a book or a paper?

۴-۳-۱۴ ساختن پاراگراف بوسیله یادداشتها

در این روش عباراتی در اختیار امتحان شونده قرار می‌گیرد تا آنها را به یک پاراگراف تبدیل کند.

مثال ۱۳

Ali/ 13 years old / study English / hobby photography / ...

۵-۳-۱۴ تشکیل پاراگراف

پاراگرافی با جملات به هم ریخته برای امتحان شوندگان در نظر گرفته می‌شود. به این ترتیب که آنها اولین جمله را به عنوان جمله موضوع انتخاب کرده و بقیه پاراگراف را با قرار دادن جملات به طور صحیح در کنار هم بنویسند. این نوع سؤال به صورت چند گزینه‌ای هم نوشته می‌شود. بدیهی است، که شکل چند گزینه‌ای کاربرد محدودتری خواهد داشت. جملات موجود در پاراگراف، به چندین روش قابل قبولی مجدداً مرتب می‌شوند. به علاوه، اگر سؤال دارای چهار گزینه باشد، پاراگراف نیز باید الزاماً شامل چهار جمله باشد.

مثال ۱۴) جملات زیر را بخوانید سپس ترکیب درست جملاتی را که پاراگراف منسجمی (خوبی) را تشکیل می‌دهند، انتخاب کنید.

- | | |
|--|---------------|
| 1. Finally, the prices are as high as in the restaurants close to the campus. | a) 2, 4, 1, 3 |
| 2. There are three reasons why I don't like To eat at the university self service. | b) 1, 2, 3, 4 |
| 3. Second, the quality of its food is poor. | c) 4, 3, 1, 2 |
| 4. First, there is no variety in the type Of food it offers. | d) 3, 2, 4, 1 |

۶-۳-۱۴ تکمیل پاراگراف

پاراگراف ناقصی در اختیار امتحان شوندگان قرار می‌گیرد، و از آنها خواسته می‌شود با افزودن یک، دو یا سه جمله از خودشان آن را کامل کنند.

مثال ۱۵

Here are four reasons why I do not like to eat at the university self – service. First, there is no variety in the type of food it offers. Second the quality of food is poor.

ع-۱۴ نوشته آزاد

نوشتن در سطوح (بالتر) متوسط و پیشرفته، اساساً به عواملی مربوط است که نیازمند تبدیل کلمات به جملات صحیح نحوی می‌باشند، که این جملات عملاً اندیشه یا عقیده‌ای خاص را منتقل می‌کنند. در نتیجه، آزمون نوشتن در این سطوح عمدتاً نگارش آزاد می‌باشد.

بدیهی است این کار ارزیابی مقاله را از آزمونهای مهارت نوشتاری جدا می‌کند. به این دلیل که مقالات، علاوه بر تواناییهای موجود در انشاء، شامل خلاقیت و ابتکار می‌باشند. از آنجائیکه این تواناییها معمولاً در شرایط امتحانی شکوفا نمی‌شوند، پس نباید آنها را در زمره آزمونهای نوشتاری قرار داد.

نوشته آزاد از نظر ظاهری، معتبرترین آزمون نوشتاری می‌باشد. انشاء نویسی بنا به دو دلیل نسبت به سایر آزمونها ارجحیت دارد. دلیل اول اینکه، انشاء نویسی به امتحان شوندگان این امکان را می‌دهد تا توانایی خود را در سازماندهی و بیان عقیده‌شان، استفاده از واژگان، سیاق و سبک مخصوص به خود نشان دهند. دوم این‌که، از آنجائیکه امتحان شوندگان عملاً می‌نویسند، آزمونهای نگارش، آنها را در بهبود نوشته‌شان فعال می‌کند. با این حساب، آزمونهای نگارش تأثیر جانبی بر آموزش نوشتن دارند. مطلب بعدی بر روشهای ارزیابی امتحان شوندگان در نوشتن انشاء تأکید دارد.

نوشته آزاد یا انشاء نویسی، اساساً، شامل موضوعی برای نوشتن انشایی با طول معین می‌باشد. امتحان شوندگان باید با محتوای موضوع آشنایی داشته باشند و همچنین راهنماییهای دقیقی در مورد موضوع خواسته شده از آنها وجود داشته باشد.

مثال ۱۶) انشایی در مورد هدف خود از خواندن زبان انگلیسی بنویسید. انشای شما بایستی در حدود ۱۵۰ کلمه باشد؛ برای انشای طولانی‌تر، نمره بیشتری داده نخواهد شد. برای کاربرد واژگان مناسب، ساختار صحیح و چگونگی بیان عقاید به بهترین نحو، نمره منظور خواهد شد. قبل از شروع به نوشتن چارچوبی برای خود بنویسید، این خلاصه به شما در سازماندهی افکارتان کمک خواهد کرد. اما این کار هیچ نمره‌ای نخواهد داشت. مدت ۳۰ دقیقه برای نوشتن انشاء در نظر گرفته شده است.

منتقدان بر این باورند که نمره‌دهی انشاء کاری ذهنی است. با این حال، می‌توان ارزیابی را با در نظر گرفتن عوامل مؤثر در عدم پایایی به نحو مؤثری به انجام رساند. برای مثال، تعیین معیار، اولین شرط می‌باشد. به علاوه، در اصلاح انشاء وجود بیش از یک مصحح، ترجیحاً سه مصحح لازم می‌باشد.

ارزیابی انشاء به دو روش صورت می‌گیرد: نمره دهی برداشتی و تحلیلی. در نمره دهی برداشتی، ممتحن به‌طور اجمالی به نوشته نگاه می‌کند، و درجات عالی / خوب / متوسط، قبول / مردود یا حتی نمره‌ای طبق سیستم مقیاس در نظر می‌گیرد تا آنجا که نمره‌دهی براساس برداشت مصحح باشد، کار

سریع و در زمان کمی انجام خواهد گرفت. برعکس در نمره‌دهی تحلیلی، امتحان‌ها یک طرح نمره‌دهی بکار می‌گیرند. هر مصححی، انشاء را به دقت خوانده و برای هر بخش نوشته، از جمله: سازماندهی، بیان منطقی عقاید، ساختار، نقطه گذاری، هجی کلمات، علامت‌گذاری، سبک و طرز بیان نمره‌ای را تعیین می‌کند. نمره انشاء مجموعه این نمرات مجزا تقسیم بر تعداد امتحانین می‌باشد. از آنجائیکه نمره‌دهی از طریق ارزیابی‌های مختلف بخشهای گوناگون نوشته صورت می‌گیرد، در واقع روش کند و وقت‌گیری می‌باشد با این وجود، این روش کارآمد بوده و بر روشهای دیگر برتری دارد.

دلایل متعددی وجود دارد که مهمترین دلیل این است که، نمره‌دهی تحلیلی، تشخیصی می‌باشد. جدول ۱-۱۴ مقیاس بارزی را برای کاربرد چنین اهدافی ارائه می‌دهد. هنوز مطالعات کاملی برای تعیین میزان نمره‌ای که در تعیین نمره کلی به هر بخش تعلق می‌گیرد صورت نگرفته است.

جدول ۱-۱۴ مقیاس درجه‌بندی انشاء

بخش	نمره	شرح اجرایی
سازماندهی	۲۰-۱۸	عنوان مناسب، پاراگراف مقدماتی به‌طور مؤثر ارائه شده است موضوع بیان شده است و به بخش اصلی مربوط است؛ عبارات انتقالی به‌کار رفته‌اند؛ ترتیب مطالب نشانگر طرح می‌باشد (طوریکه خواننده می‌تواند چارچوب بندی کند) شواهد پشتیبان برای نتایج کلی؛ نتیجه‌گیری کامل و منطقی.
مقدمه بخش اصلی و نتیجه‌گیری	۱۵-۱۷	عنوان مناسب، مقدمه و نتیجه‌گیری کافی بوده و بخش اصلی قابل قبول می‌باشد، برخی عوامل وجود ندارند. بعضی عقاید به‌طور کامل ارائه نشده‌اند؛ توالی مطالب منطقی می‌باشد، ولی عبارات انتقالی یا وجود ندارند یا درست به‌کار نرفته‌اند.
	۱۲-۱۴	مقدمه و نتیجه‌گیری متوسط یا ناقص می‌باشد؛ در نظم مطالب موجود در بخش اصلی مشکلاتی وجود دارد؛ نتیجه کلی به‌طور کاملی با دلایل ارائه شده تأیید نشده است؛ مشکل سازماندهی وجود دارد.
	۶-۱۱	مقدمه بی‌ثبات و در حد پائین می‌باشد. سازماندهی کمی وجود دارد. مشکلات اساسی در نظام عقاید وجود دارد. فاقد دلایل تأییدی می‌باشد نتیجه‌گیری ضعیف و غیرمنطقی می‌باشد. سازماندهی کافی نمی‌باشد.
	۱-۵	مقدمه یا نتیجه‌گیری وجود ندارد؛ ظاهراً هیچ سازماندهی در بخش اصلی به‌چشم نمی‌خورد؛ شواهد تأییدی اصلاً وجود ندارند، نویسنده برای سازماندهی انشاء هیچ اقدامی نکرده است (خواننده نمی‌تواند چارچوب متن را تصور کند).

مقاله به موضوع تعیین شده مربوط است؛ مطالب ملموس و بطور کامل مبسوط می‌باشند؛ مطالب اضافی وجود ندارد؛ مقاله بازتاب عقاید می‌باشد.	۱۸-۲۰	ارائه منطقی مطالب: محتوی
مقاله خطاب به مطالب اصلی می‌باشد. اما به برخی نکات اشاره‌ای نشده است؛ عقاید باید کاملتر ارائه می‌شدند؛ برخی مطالب اضافی وجود دارند.	۱۵-۱۷	
گسترش مطالب کامل نمی‌باشند یا مقاله تا حدی خارج از موضوع می‌باشد؛ پاراگراف‌ها به‌طور دقیق جدا نشده‌اند.	۱۲-۱۴	
مطالب ناقص بوده و مقاله بازتاب تفکر دقیق نمی‌باشد و با عجله نوشته شده است؛ در مورد محتوا اقدام کاملی صورت نگرفته است.	۶-۱۱	
مقاله کاملاً نامناسب بوده و نشان‌دهنده کاری در سطح دانشگاهی نمی‌باشد؛ و موضوع به دقت مورد توجه قرار نگرفته است.	۱-۵	
در ساختار متن انگلیسی، فصاحت زبان اصلی وجود دارد؛ شبه جمله‌ها حروف اضافه، افعال کمکی، حروف تعریف، فرم افعال و توالی زمانها به‌طور صحیحی به‌کار رفته‌اند؛ گسیختگی یا ادغام در جملات وجود ندارد.	۱۸-۲۰	ساختار
مهارت پیشرفته در ساختار انگلیسی؛ وجود برخی اشکالات ساختاری که در برقراری ارتباط مطالب تأثیری ندارد، اگر چه خواننده از آنها مطلع می‌باشد؛ گسیختگی یا ادغام در جملات وجود ندارد.	۱۵-۱۷	
مطالب برای خواننده قابل فهم می‌باشد، اما اشکالات ساختاری موجود تأثیر منفی در انتقال مطالب دارند؛ گسیختگی یا ادغام در جملات به چشم می‌خورد.	۱۲-۱۴	
اشکالات ساختاری متعدد در انتقال عقاید نویسنده تأثیر دارند؛ ساختار از برخی جهات نیاز به مرور دارند؛ خواندن جملات مشکل می‌باشد.	۶-۱۱	
اشکالات اساسی ساختاری در انتقال پیام تداخل ایجاد می‌کنند؛ خواننده منظور نویسنده را نمی‌فهمد، ساختار جملات غیرقابل فهم می‌باشد.	۱-۵	
کاربرد صحیح قوانین نوشتاری انگلیسی، حاشیه‌گذاری از هر دو طرف، نوشتن حروف بزرگ لازم، فاصله پاراگرافها، و هجی صحیح کلمات.	۱۸-۲۰	نقطه‌گذاری هجی کلمات نکات تخصصی

- ۱۵-۱۷ برخی اشکالات در قوانین نوشتاری یا نقطه‌گذاری؛ اشکال در هجی کلمات، حاشیه‌گذاری صحیح در سمت چپ؛ تکلیف تمیز و خوانا است.
- ۱۲-۱۴ مقررات کلی نوشتاری به کار رفته است، اما اشکالاتی وجود دارد؛ اشکالات املایی توجه خواننده را منحرف می‌کند؛ و اشتباهات نقطه‌گذاری بیان عقاید را مختل می‌کند.
- ۶-۱۱ اشکالات اساسی در فرم تکلیف وجود دارد؛ بخشهایی از مقاله ناخوانا می‌باشد؛ اشتباهاتی در نقطه‌گذاری نهایی جمله وجود دارند که برای خوانندگان باسواد غیرقابل قبول می‌باشند.
- ۱-۵ به قوانین نوشتاری انگلیسی توجهی نشده است؛ تکلیف ناخوانا بوده و حروف الفبای بزرگ به کار نرفته است، حاشیه وجود ندارد، و اشکالات املایی مهمی به چشم می‌خورد.
- ۱۸-۲۰ سبک و طرز بیان کاربرد دقیق لغات و استفاده از ساختارهای همسان، نوشته به‌طور خلاصه، سیاق مناسب.
- ۱۵-۱۷ وجود تنوع، لغات خوب و نه زیاد پر طول و تفصیل، سیاق مناسب، سبک نسبتاً مختصر.
- ۱۲-۱۴ کاربرد غلط برخی لغات، عدم توجه به سیاق، تقریباً پر طول و تفصیل.
- ۶-۱۱ بیان ضعیف عقاید، اشکال در کاربرد لغات، عدم تنوع در ساختار.
- ۱-۵ کاربرد نامناسب لغات، عدم وجود مفهوم، سیاق یا تنوع در جمله.
- (Bailey and Brown, 1984)

۵-۱۴ دستورالعملهایی برای آماده کردن آزمونهای نوشتاری

برای این که آزمونی دارای اعتبار و پایایی باشد، ممتحن باید در ساختن سؤالات و نمره‌دهی به پاسخهای امتحان‌شوندگان توجه زیادی کند. معتبرترین نوع آزمون نوشتاری آزمونی است که امتحان‌شونده را ملزم به نوشتن کند. در حالی که آزمونهایی که به‌طور عینی نمره‌دهی می‌شوند به امتحان‌کننده در ارزیابی مطالب مربوط بسیار کمک می‌کند و نوشته آزاد به امتحان‌شونده این امکان را می‌دهد تا از ذکر نکاتی که مطمئن نیست، اجتناب کند. اگر اهداف امتحان از لحاظ هزینه و افراد تأمین شوند، ایده‌آل‌ترین آزمون نوشتاری، نوشته آزاد می‌باشد. نکته مهم دوم، ترکیب توأم سؤالهایی که به طریق عینی نمره‌دهی می‌شوند و نوشته آزاد، معیار بهتری در مورد مهارتهای نوشتن نسبت به تک تک سؤالها به دست می‌دهد. اما اگر لازم باشد مهارت نوشتاری عده زیادی مورد ارزیابی قرار گیرد،

سؤالاتی که به روش عینی نمره‌دهی می‌شوند مناسب‌تر خواهد بود. مطالب بعدی، نظریات سودمند، در طرح و نمره‌دهی انشاءها ارائه می‌دهد.

۱- تکالیف موجود در آزمونهای نوشتاری باید نشانگر انواع مختلف کلام و محتوا باشند. طراح آزمون باید موقعیتهای نوشتاری واقعی که امتحان شوندگان آن را عملاً تجربه کرده یا احتمالاً در آینده با آن مواجه خواهند شد، مورد توجه قرار دهد. نوشتن نامه‌ها، یادداشتها، خاطره‌ها، گزارشها و امثال اینها بیش از موضوعات انتزاعی به واقعیتهای زندگی نزدیکتر می‌باشند. باید از موضوعات بیهوده چون «چشمهای او» «ثروت یا دوستی» اجتناب کرد. موضوعاتی که برای کارهای نوشتاری انتخاب می‌شوند، باید دارای هدف بوده و در امتحان شونده، انگیزه‌ای برای بیان مطلب به خواننده‌ای فرضی ایجاد کند.

برای مثال یک از دوستانتان در نظر دارد خوابگاه را ترک کند و آپارتمانی را بیرون از دانشگاه اجاره کند، اما نمی‌تواند تصمیم بگیرد. او نظر شما را می‌پرسد نامه‌ای به او بنویسید و مزایا و معایب هر کدام را بنویسید.

چنین کاری بیش از پرسش ساده در مورد «مزایا و معایب زندگی در محوطه دانشگاه ایجاد انگیزه می‌کند. کارهای نوشتاری که براساس اطلاعات موجود در تصاویر، پوسترها، آگهی‌ها، جدول‌ها و نمودارهای محیط پیرامون و غیره می‌باشند، کاملاً روشن‌گر بوده و انگیزه کاملی در فرد ایجاد می‌کند.

۲- هر یک از پرسشهای آزمون باید چنان نوشته شده باشند که تمام امتحان شوندگان همانند ممتحنین درک یکسانی از آن داشته باشند، نباید از امتحان شوندگان انتظار داشت تا آنچه را که از آنها خواسته می‌شود حدس بزنند. طرح موضوعاتی همچون «پول» و «اقتصاد» کار صحیحی نمی‌باشد. شاید امتحان شونده‌ای مفصلاً در مورد روش به‌دست آوردن پول و دیگری در مورد مزایایی پولداری بنویسید در حالی که شاید فرد دیگر پولداری را با دانش اندوزی مقایسه کند. بهترین تکلیف، تکلیف مشخص همانند نمونه زیر می‌باشد:

به‌دست آوردن پول کار مشکلی می‌باشد، اما خرج کردن آن آسان است. اگر بیش از نیازتان پول داشتید، چه کار می‌کردید؟

۳- کار نوشتاری باید برای تمام امتحان شوندگان یکسان باشد، نباید حق انتخاب وجود داشته باشد. زیرا در این صورت، مقایسه بین امتحان شوندگان اگر غیرممکن نباشد، ولی بسیار مشکل خواهد بود و نمره‌دهی به میزان پایا نخواهد بود. به علاوه، تصمیم‌گیری در مورد انتخاب یک موضوع برای امتحان شونده کار سختی خواهد بود.

۴- انشاءهای طولانی برای امتحان شونده خسته کننده بوده و نشان دهنده توانایی آنها نخواهد بود. به علاوه شاید تمام امتحان شوندگان به موضوع خاصی علاقه نداشته باشند. از دو انشاء کوتاه نتایج پایا و معتبری به‌دست می‌آید.

۵- نباید ممتحن بگذارد نمراتش تحت تأثیر هنرگرایی نامربوط قرار گیرد. برای این که نمره دهی عینی باشد، برای مثال تمیزی نباید در تصحیح انشاء تأثیر گذارد.

ع باید هر انشایی بدون اشاره به نویسنده‌اش تصحیح شود. این نوع نمره دهی اساس تجربه یا شناخت قبلی را کاهش می‌دهد.

۷- اگر آزمونی از بیش از یک انشاء تشکیل شده باشد، یکی از انشاهای تمام امتحان شوندگان تصحیح شود و سپس ممتحن به سراغ تکالیف بعدی برود. این امر قضاوت یکسانی را موجب می‌شود به علاوه عملکرد یک امتحان شونده در یک انشاء در نمره‌دهی انشاء دیگر همان فرد تأثیری نخواهد داشت.

فعالیتها

- ۱- معایب و مزایای نمره‌دهی برداشتی را ذکر کنید.
- ۲- بخشهای زیر، برخی روشهای آزمون مهارتهای نوشتاری را نشان می‌دهد. نظر شما در مورد ارزیابی هر یک از سؤالات چیست؟ مزایا و معایب هر کدام را ذکر کنید.
- الف) برای مثال آزمونی از امتحان شوندگان می‌خواهد با استفاده از سبک عامیانه، غیررسمی یا رسمی، متنی را دوباره بازنویسی کنند. کار آنها براساس روش انتقال بهتر متن، ارزیابی خواهد شد.
- ب) روش دیگر استفاده از ترجمه می‌باشد. از امتحان شوندگان خواسته می‌شود متنی را از زبان خود به انگلیسی ترجمه کنند.
- ج) در روش دیگر از امتحان شوندگان خواسته می‌شود تا متنی را بخوانند و به سؤالاتی پاسخ دهند. آنها باید در قالب جملات ساده به سؤالات پاسخ دهند به این ترتیب که آنها را به صورت متنی منسجم با توالی منطقی بنویسند. پاسخ سؤالات پاراگرافی را تشکیل می‌دهند که خلاصه‌ای از متن باشد.
- چ) بالاخره علاوه بر دو روش گفته شده روش دیگری برای نمره‌دهی انشاء، شمارش اشتباهات می‌باشد. این کار شامل شمارش خطاهای امتحان شوندگان و کم کردن آن از نمره کلی می‌باشد. برای مثال شاید امتحان شونده‌ای نمراتی را به خاطر خطاهای ساختاری، کاربرد نادرست لغات و یا املای غلط از دست بدهد.

۳- محققان معتقدند که همبستگی بالایی میان نمره‌دهی برداشتی و تحلیلی در تعداد زیادی از انشاها وجود دارد. تصور کنید، نمرات به دست آمده از طریق دو شیوه نمره‌دهی در مورد یک موضوع را باهم مقایسه می‌کنیم. آیا نتایج به دست آمده یکسان خواهد بود، یا متفاوت؟ از این مشاهده چه نتیجه‌ای می‌گیرید؟

۴- آیا همه ممتحنین استدلال یکسانی از نمره‌دهی برداشتی خواهند داشت؟ با دو معلم که چنین قضاوتی دارند مصاحبه کرده و نتیجه کار را باهم مقایسه کنید.

۵- نتیجه تحقیق نشان‌دهنده وجود همبستگی بالایی میان آزمونهای نوشته آزاد و آزمونهای نوشتاری که به طریق عینی نمره‌دهی می‌شوند، می‌باشد. این جمله به چه معناست؟ شما از کدام روش برای ارزیابی نوشته استفاده می‌کنید؟ کدام عوامل را مورد توجه قرار می‌دهید؟

آزمونهای جای خالی و املاء

۱۵-۱ مقدمه

در فصول قبل، بخش‌ها و مهارت‌های زبانی آزمون‌سازی بطور جداگانه مورد بررسی قرار گرفتند. برای ارزیابی این بخش‌ها و مهارت‌ها روش‌های گوناگونی ارائه شدند. تجزیه زبان به مهارت‌های زبانی و اجزاء سازنده آن از پیدایش و حاکمیت روش آزمون مجزا شکل، ناشی می‌شود. بنابه دلایلی این روش بیش از ربع قرن از اعتبار بالایی برخوردار بود. در بیشتر نقاط دنیا بنابه دلایلی این روش یکی از روش‌های معتبر برای ارزیابی مهارت‌های زبانی به حساب می‌آید.

اولین دلیل این بود که آزمونهای مجزا شکل می‌توانستند به اهداف متعالی دست یابند. برای مثال، این آزمون‌ها به آسانی برای اهداف تشخیصی به کار می‌رفتند و در برنامه‌های آموزشی در طرح مطالب جبرانی مورد استفاده قرار می‌گرفتند. مثال دیگر آنکه آزمونهای مجزا شکل اطلاعات پایایی در مورد تواناییهای امتحان‌شوندگان در هر یک از زمینه‌های زبانی فراهم می‌کردند. برخی محققان براین باور بودند که هر یک از فراگیران زبان تواناییهای متفاوتی در زمینه‌های مختلف زبان دارند، اطلاعات به‌دست آمده از طریق آزمونهای مجزا شکل می‌توانست به مربیان کمک کند تا برنامه‌های مفیدی را برای برطرف کردن این تفاوتها طراحی کنند.

دلیل دوم آن بود که آزمونهای مجزا شکل دارای ویژگیهای روانسنجی مطلوبی بودند. مطالعات متعددی نشان دهنده ارقام معتبر و پایایی بالایی برای این نوع آزمون‌ها می‌باشند. با توجه به ارزیابی هر چه علمی‌تر فعالیتهای سنجش، ممتحنین از ویژگیهای علمی این نوع آزمون‌ها استقبال کردند.

دلیل سوم این بود که آزمونهای مجزا شکل کاملاً کاربردی بودند. به منظور ارزیابی توانایی درک مطلب فراگیران از طریق آزمونهای چند گزینه‌ای، اجرای آزمون و فرآیندهای نمره‌دهی تا حد زیادی تسهیل یافت. مخصوصاً با بکارگیری پیشرفتهای وابسته به فن‌آوری از طریق دستگاهها سوالات چند گزینه‌ای قابل نمره‌دهی را پیدا کردند. و این امر یکی از مزایای بزرگ آزمونهای مجزا شکل نسبت به سایر آزمونها بود.

علیرغم این مزیتها، آزمونهای مجزا شکل بنابه برخی معایب نظری نمی‌توانستند بر سایر آزمونها برتری داشته باشند. این دلایل تا حدی از پیشرفت زبانشناسی و تا اندازه‌ای از پیشرفت در زمینه روانشناسی و روانسنجی ناشی می‌شد.

در وحله اول، پیدایش نظریه زایاگشتاری به‌طور جدی اعتبار زبان را از نظر ساختاری، یادگیری زبان و آزمون سازی زبان زیر سؤال برد. بنابراین فرضیه تقسیم زبان به اجزاء و قسمتهای مجزا بدون توجه به بافتی که این اجزا در آن به کار می‌روند، دیگر قابل قبول نبود. پیشرفتهای اخیر ثابت کرده است که واجها به تنهایی بدون اشاره به تکواژها و تکواژها به جملات و جملات به بافتهای بزرگتر وجود نخواهند داشت یا اگر وجود داشته باشند، بی‌فایده خواهند بود.

به عبارت دیگر، تئوری آزمون مجزا شکل این فرضیه را ارائه می‌دهد که زبان قابل تقسیم به عناصر مجزا می‌باشد و می‌توان این عناصر را جداگانه مورد ارزیابی قرار داد. این گونه بحث می‌شد که در فرآیند آموزش می‌توان بر اجزاء سازنده مهارتی در بافت اجزاء یا مهارتهای دیگر تمرکز کرد. اما جداسازی کامل یا تفکیک سازه‌ها و مهارتها به منظور آزمون، امری غیرممکن به نظر می‌رسد. واج‌شناسی، ساختارشناسی، نحو و معناشناسی - اجزاء زبان - در درک یا بیان مهارت زبانی نقش دارند. این مهارتها همچنین در درک و بیان یکدیگر نیز تأثیر دارند. نمودار ۱-۱۵ رابطه بین سازه‌های زبان و مهارتهایی که شبکه ارتباط زبانی را تشکیل می‌دهند، نشان می‌دهد.

شکل ۱-۱۵- شبکه زبانی

COMPONENTS	SKILIS	MODE & PROCESS
Phonology	Listening	Productive
Morphology	Speaking	Oral
Syntax	Reading	Written
Semantics	writing	Receptive

از نمودار بالا چنین برمی‌آید که در شبکه زبانی هیچ بخش آزادی وجود ندارد. یعنی هیچ کدام از سازه‌ها یا مهارتها مستقل عمل نمی‌کنند. همه آنها در این شبکه برای این که زبان به عنوان ابزاری ارتباطی عمل کند باهم در ارتباطند.

در مرحله دوم، طرفداران آموزش ادراکی (شناختاری)، بر این باورند که حتی اگر زبان قابل تجزیه به عناصر مجزا بود، نقش هر عنصر در کاربرد کل زبان معلوم نمی‌شد و همچنین چنین نقشی در

جریان کلی ارتباط، اهمیتی نداشت. بنابراین نمرات یا مجموعه آنها برحسب این عناصر، شاخص معتبری از توانایی فرد نبود.

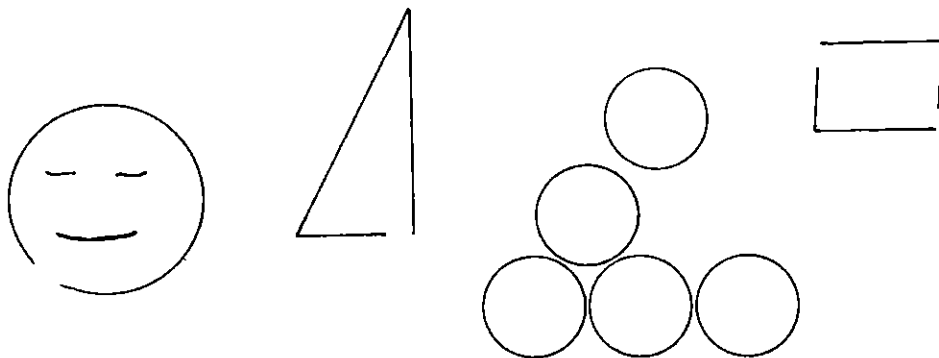
در وحله سوم، با گسترش روانشناسی گشتالت، روشی را که فرآیندهای زبانی و یادگیری زبان توصیف شده بود، تحت تأثیر قرار گرفت. طبق روانشناسی گشتالت، کل، چیزی متفاوت از مجموعه اجزای تشکیل دهنده است. با تعمیم این اصل به جریان فرآیندهای زبان دانشمندان پذیرفتند که کل توانایی زبان از مجموعه تواناییهای مجزا متفاوت است. به عبارت دیگر، عناصر زبانی افزایشی نمی باشند. بحث های فوق الذکر در مورد آزمونهای مجزا شکل منجر شد که حوزه آزمون سازی به دنبال رویکرد دیگری باشد. محققان اقدام به طرح آزمونهایی کردند، که از طریق آنها زبان به عنوان پدیده ای کلی مورد ارزیابی قرار می گرفت. این حرکت در زمینه آزمون سازی به پیدایش روش تلفیقی انجامید از میان آزمونهای متداول که تواناییهای زبانی را به صورت کلی ارزیابی می کنند، می توان آزمونهای جای خالی و املاء را نام برد. بنابراین در ادامه این فصل به توضیح در مورد این دو آزمون خواهیم پرداخت.

۲-۱۵ فرآیند جا خالی

ریشه کلمه cloze به سال ۱۹۵۳ برمی گردد. یعنی زمانیکه تیلور^۱ در تلاش برای طرح مطالب خواندنی مناسبی برای دانش آموزان بومی زبان بود. وی کلمه cloze را به عنوان یادآوری از مفهوم closure که در روانشناسی گشتالت مفهومی شناخته شده بود، ابداع کرد. روانشناسی گشتالت، که تا اواخر قرن نوزدهم در اتریش و آلمان توسعه یافته، به ویژه در زمینه درک تجسمی موثر بود. طرفداران این مکتب روانشناسی از جمله ورتیمر (۱۹۴۳ - ۱۸۸۰) کوهلر (۱۹۶۷-۱۸۸۷) و کافکا (۱۹۴۱-۱۹۸۹) (به اهمیت نظری الگوی محرک در ایجاد تجربه ادراکی پی بردند. گشتالت واژه ای آلمانی است که معادل دقیقی در زبان انگلیسی ندارد. این واژه به مفهوم تاثیر ویژگیهای کلی بر روش درک اجزا تاکید دارد، ادراک سعی می کند تا داده های حسی را در یک الگوی کلی یا گشتالتی متمرکز کند، به همین دلیل گاهی این تئوری را بعنوان تئوری کل نیز می نامند. فرضیه اصلی روانشناسی گشتالت بر این اصل استوار است که کل چیزی متفاوت از مجموعه اجزاست. طرفداران این نظریه عقیده ساخته شدن ادراک از اجزای متمایز حس را رد می کنند.

نتیجه ای که از تئوری استنباط می شود این است که انسان توانایی درک اشکال ناتمام و ناقص را به صورت اجسام کامل دارد. مثالی در مورد درک تجسمی، تکمیل اشکال هندسی ناقص، شکلی با حالت بسته می باشد، مانند مثالی که در زیر آمده است:

^۱. Tylor



بنابراین اگر فردی شی‌ای را تجسم کند، این شی حتی اگر دارای قسمت‌های ناقص و یا خالی باشد، به‌طور کامل درک خواهد شد. با تعمیم این نگرش به فرآیند زبانی، می‌توان گفت که اگر فردی مهارت کاربرد زبان را در خود ایجاد کرده باشد، او خواهد توانست درک خود را از یک بخش زبان کامل کند، حتی اگر آن بخش از زبان ناقص یا جاهای خالی داشته باشد.

توانایی پر کردن جاهای خالی متن، نشانگر توانایی درک مطلب خواننده‌ها می‌باشد. بنابراین، روش جای خالی، تبدیل به شاخص قابلیت خواندن در زبان اول شد. در واقع تیلور روش جا خالی را بعنوان جایگزینی برای آزمونهای خواندن و درک مطلب پیشنهاد کرد. او معتقد بود هنگامیکه جا خالی‌های خاصی، یعنی جاهای خالی در متن ایجاد شود، باعث کلیت ناقص می‌شود. کلمات اطراف جای خالی به‌عنوان محرکی برای فعال کردن مکانیسم ذهنی خواننده عمل می‌کنند. بنابراین خواننده باید تلاش کند تا اطلاعاتی از متن بیابد که در پیدا کردن لغات حذف شده کمک کنند.

در این حالت، توانایی خواننده در تکمیل متن ناقص بطور کلی نشانگر توانایی خواندن و درک مطلب و بخصوص توانایی پردازش زبان خواهد بود. در اصطلاح تخصصی‌تر، فرآیندهای جاخالی وسیله‌ای برای ارزیابی دانش دستوری درونی شده فراگیران می‌باشد. در حقیقت فرآیند جاخالی نشانگر کارایی فراگیران در کاربرد فرآیندهای دستوری به منظور تکمیل بخشهای حذف شده و یا ناقص متن می‌باشد. برنامه‌های تحقیقی متعددی برای بررسی کارایی فرآیند جاخالی به‌عنوان آزمونی برای ارزیابی مهارت زبان آموزان غیربومی زبان صورت گرفته است. براساس این برنامه‌ها، مفهوم جاخالی با ویژگیهای دقیق‌تر و علمی‌تر تثبیت شده است.

طبق نظریه تیلور، فرآیند جاخالی چنین تعریف می‌شود: هر متنی با طول و سطح دشواری مناسب با کلمه n حذف شده، تعدادی از اصطلاحات این تعریف نیاز به توضیح دارند. منظور از طول مناسب چیست؟ سطح دشواری مناسب به چه معناست؟ مقدار n چیست؟ برای پاسخ به این سؤالات تحقیقات فراوانی به وقوع پیوست.

اولین بخش تحقیق در مورد تعیین میزان n می‌باشد. واضح است که حداقل میزان n «۲» است. یعنی کلمات به صورت یک در میان باید حذف شوند. اگر چه تعیین حداقل میزان آسان است، اما در مورد حداکثر چنین نمی‌باشند. البته چنانچه میزان n افزایش یابد، درک متن و یافتن کلمات حذف شده آسانتر خواهد شد، زیرا بافت بیشتری در اختیار خواننده قرار خواهد گرفت.

برای تعیین میزان مناسب n آزمونهای جاخلی با حذف کلمه سوم، چهارم، پنجم، ششم، هفتم، هشتم، نهم و دهم طرح شدند. در عمل، آزمونهایی با هر پنجمین یا هفتمین کلمه حذف شده مناسب‌تر از همه بودند. از نظر تجربی نیز، آزمونهایی با این میزان حذف، ویژگیهای روانسنجی بهتری از خود نشان داده‌اند. خصوصاً آزمونهای جا خالی با حذف هر هفتمین کلمه بعنوان آزمون جاخلی استاندارد شناخته شده می‌باشند. اگر چه نسبت به سایر آزمون‌های جاخلی (با نسبتهای مختلف) از مزیت بیشتری برخوردار نبود. اما به عنوان تستهای جاخلی استاندارد شناخته شدند.

دومین بخش تحقیق در مورد تعیین تعداد مناسب کلمات حذف شده بین ۲۰ تا ۱۰۰ عدد بود. از آنجائیکه تعداد کلمات حذف شده طول متن را تعیین می‌کنند، دقت بیشتری در مورد آنها به عمل آمد. با توجه به روشهای قدیمی در طرح آزمونهای درک مطلب اکثر محققان معتقدند که طول متن نباید بیشتر از ۳۰۰-۴۰۰ کلمه باشد. با این تعداد کلمه امکان حذف ۴۰-۵۰ کلمه وجود دارد. اما، رند (۱۹۷۸) یک برنامه تحلیل پرسشی کامپیوتری طراحی کرد که به‌طور تصنعی آزمونهای جاخلی را در هر نوبت یک پرسش بیشتر می‌کرد. این برنامه آماری برای آزمونی با دو، سه، چهار ... تا پنجاه پرسش ایجاد کرده بود. تا حدود بیست و پنج سؤال بیشترین پایایی به دست می‌آمد. نتیجه برنامه این بود که با افزایش تعداد سؤالات به بیش از ۲۵ عدد دقت آزمون کمتر می‌شد. اگر چه یک آزمون ۲۵ سؤالی، میزان پایایی قابل قبولی را نشان می‌دهد، اما از اعتبار بالایی برخوردار نمی‌باشد. بنابراین تعداد پرسشها باید به‌گونه‌ای تعیین شود که پایایی و اعتبار بالایی به دست آید. تجربیات چنین نشان می‌دهد که آزمونهای جاخلی با ۲۵-۳۰ پرسش شاخص اعتبار بالایی را همچون آزمونهای ۴۰-۵۰ سؤالی نشان می‌دهند. بنابراین، معیار صحیح تعداد حذف در آزمون جاخلی حدود ۲۵-۳۰ عدد تعیین شده است، به این معنا که با در نظر گرفتن میزان حذف هفتمین کلمه و ۲۵ الی ۳۰ جای خالی، طول متن چیزی بین ۲۱۰ الی ۱۷۵ کلمه خواهد شد.

مطلب دیگری در مورد طول متن جاخلی پیش کشیده شد. با توجه به ساختار پاراگراف در زبان انگلیسی، محققان براین باورند که یک پاراگراف معمولی شامل یک جمله بنیادی، جملات کمکی یا پشتیبان و جمله نتیجه‌گیری می‌باشد. در واقع، معمولترین نوع پاراگراف، اصطلاحاً پاراگراف I فرم نامیده می‌شود.

یعنی، جمله بنیادی، که بیشترین بار اطلاعاتی را در خود دارد، معمولاً در آغاز پاراگراف قرار می‌گیرد، و جمله نتیجه‌گیری که خلاصه‌ای از محتوای پاراگراف است، در پایان می‌آید. بنابراین، چنین برمی‌آید که حذف کلمه‌ای از این جملات، درک مطلب امتحان شونده را مختل کند. به عبارت دیگر،

این دو جمله، بیشترین مقدار اطلاعات را که نقش کلیدی را در فهم کل مطلب دارند، فراهم می‌کند. بنابراین، تصور می‌شد که آوردن اولین و آخرین جمله بدون حذف کلمه‌ای به درک آسانتر خواننده از متن کمک خواهد کرد، و بالاخره آوردن اولین و آخرین جمله بدون حذف کلمه‌ای به صورت یک اصل پذیرفته شده درآمد. در واقع تحقیق بعدی این فرضیه را تایید کرد. با اضافه کردن تعداد کلمات در این دو جمله، متن جاخلالی به ۲۵۰ - ۲۲۰ کلمه افزایش خواهد یافت. با توجه به نتایج اولین و دومین بخش تحقیق با در نظر گرفتن تعداد کلمات بین دو حذف و تعداد حذفها، تعریف دقیق‌تری از آزمون جاخلالی به وجود می‌آید: آزمون جاخلالی متنی است با سطح دشواری مناسب و حدود ۲۵۰-۲۲۰ کلمه با حذف هر هفتمین کلمه چیزی که باقی می‌ماند تعیین سطح دشواری مناسب متن می‌باشد.

سومین بخش تحقیق در مورد تعیین تقریبی سطح دشواری متن بود. آثار موجود در مورد تعیین سطح دشواری متن قدیمی و غنی می‌باشند. طی چند دهه گذشته، فرمولهای قابلیت خواندن زیادی ارائه شده است. اگرچه در ابتدا قرار بود خود فرآیند جاخلالی سطح دشواری مناسب برای خوانندگان در سطوح مختلف توانایی ایجاد کند، اما این کار بسیار مشکل بود. بنابراین محققان به کاربرد فرمولهای قابلیت خواندن برای تعیین تناسب متون جاخلالی برای افراد موردنظر رو آوردند.

اگر چه فرمولهای قابلیت خواندن در ظاهر متفاوت می‌باشند، اما از ویژگیهای مشترکی برخوردارند. اکثر آنها عواملی همچون تعداد هجاها یا کلمات، طول جملات، پیچیدگیهای نحوی جملات را به کار می‌برند. برخی از آنها مشکل و برخی آسان و عملی می‌باشند، اما این فرمولهای مختلف، متن را به یک شیوه ارزیابی می‌کنند. در واقع فرقی نمی‌کند کدام فرمول مورد استفاده قرار می‌گیرد، به این شرط که فرمول واحدی پیوسته به کار رود، (میرزایی ۱۹۹۱) یکی از فرمولهای عملی که فاگ ارائه داده است و شاخص قابلیت خواندن فاگ نامیده می‌شود به این قرار است:

(تعداد کلمات سه سیلابی) / تعداد کلمات

$$۴۰ = \frac{\text{تعداد کلمات سه سیلابی}}{\text{تعداد کلمات}} \quad \text{سطح دشواری متن}$$

تعداد کلمات / تعداد جملات

روش دیگر برآورد قابلیت خواندن متن نمودار گسترده فرای^۱ می‌باشد که در ضمیمه نشان داده شده است.

همانطور که قبلاً ذکر شد اکثر این فرمولها از پارامترهای رو ساخت استفاده می‌کنند. علاوه بر آن، در باب اعتبار این فرمولها اختلاف نظرهایی وجود دارد. همچنین این فرمولها برای گویشگران بومی زبان انگلیسی طراحی شده‌اند. بنابراین، جز در مواردی که تحقیق گسترده‌ای بعمل آید، و گرنه باید با احتیاط کامل استفاده شوند. اما تحقیق در زبان خارجی و یا زبان دوم نشان‌دهنده این است، که نمره آزمون جای خالی ۰/۵۳ یا بالاتر با آنچه که به طور سنتی سطح مستقل خواندن نامیده شده است مطابقت

^۱. Fry

می‌کند. نمره ۰/۴۴ تا ۰/۵۳ در میزان مناسب برای مطالب ابزاری، یا اصطلاحاً سطح ابزاری می‌باشد، و نمره زیر ۰/۴۴ در سطح ناموفق خواندن قرار می‌گیرد. (اولیر ۱۹۸۰)

با در نظر گرفتن پیچیدگیها و اختلاف نظرهای موجود در فرمولهای قابلیت خواندن و بررسی تفاوت‌های میان گویشگران غیربومی زبان به سختی می‌توان فرمول یا معیار خاصی را به انتخاب متن جای خالی برای گویشگران غیربومی اختصاص داد. اما روش زیر در این مورد شاید مفید باشد.

در انتخاب متنی برای گویشوران غیربومی زبان، توصیه می‌شود ابتدا محتوای مطالبی که توسط افراد موردنظر برای کسانی که آزمون بر روی آنها انجام خواهد شد (مطالعه شده یا قرار است مطالعه شوند) تعیین گردد. سپس باید نمونه تصادفی متون از محتوا انتخاب شوند. سپس فرمول خاص قابلیت خواندن در مورد این متون به کار رود و میانگین شاخص قابلیت خواندن به عنوان معیاری برای انتخاب متن نادیده برای افراد خواهد بود.

برای مثال فرض کنید قرار است آزمون جای خالی برای دانش‌آموزان کلاس چهارم دبیرستان طرح شود. ابتدا متنی از کتابشان به طور تصادفی انتخاب می‌شود و سپس یکی از فرمولهای قابلیت خواندن به کار می‌رود و میانگین قابلیت خواندن متن تعیین می‌گردد. و سرانجام متن نادیده که برای گسترش طرح آزمون جای خالی به کار خواهد رفت با همان فرمول امتحان می‌شود. معمولاً متنهایی با سطوح قابلیت خواندن انحراف معیار ± 1 ، برای این منظور مناسب می‌باشند.

به عنوان یک مثال عددی، تصور کنید پنج متن از کتاب کلاس چهارم انتخاب شده‌اند. با استفاده از شاخص فاگ سطوح قابلیت خواندن متن‌ها به ترتیب ۷، ۸، ۹، ۱۰ و ۶ محاسبه شده است. میانگین قابلیت خواندن ۸ و انحراف معیار تقریباً ۱/۵ خواهد بود. پس متون انتخاب شده برای آزمون جاخلی باید بین $8 \pm 1/5$ از میانگین، یعنی بین ۶/۵ تا ۹/۵ باشد، که با همان فرمول‌ها تعیین می‌شود. با استفاده از این روش انتخاب متنی برای گویشگران غیربومی زبان، آسان خواهد بود. علاوه نباید انتخاب مقایسه بین گویشوران بومی و غیربومی باشد، و نباید به طبقه‌بندی ذهنی امتحان شوندگان به سطوح ابتدایی متوسط و پیشرفته بستگی داشته باشد.

با تعیین سطح دشواری یک متن، تعریف واژه جاخلی که در آغاز این فصل اشاره شد به این ترتیب اصلاح می‌شود. متنی با سطح دشواری مناسب (با فرمولهای قابلیت خواندن تعیین شده است) و با طول مناسب (حدود ۲۵۰-۲۲۰ کلمه) و با حذف هر هفتمین کلمه.

این تعریف اگر چه تعریف استاندارد نیست و به طور تجربی بر سایر تعاریف ارجحیت ندارد، اما تعریفی منطقی برای کسانی است که علاقمند به کار فرآیند جاخلی می‌باشند. البته اقسام دیگر این آزمون‌ها نیز وجود دارند که ویژگیهای خوبی همانند این نمونه دارند. انواع دیگر این آزمون در زیر به تفصیل ذکر شده‌اند.

۳-۱۵ گونه‌های فرآیند جاخالی

روشهایی که برای طرح آزمون جاخالی ذکر شد، به‌عنوان نسبت ثابتی شناخته شده‌اند. یعنی میزان حذف هر پنجمین یا هفتمین کلمه ثابت است. اگر چه این روش متداولترین و معمولترین روش می‌باشد اما تحقیقات متعددی برای ایجاد تنوع در روشهای آزمون جاخالی برای اهداف مختلفی صورت گرفته است.

یکی از این تحقیقات که برای طرح آزمونهای جاخالی متفاوت از آزمون به اصطلاح جاخالی استاندارد صورت گرفته است، بر ماهیت فرآیند روانشناسی موجود در انجام آزمون جاخالی تأکید دارد. این نوع آزمون که کاملاً متداول شده است، فرم چندگزینه‌ای آزمون جاخالی می‌باشد. در یک آزمون جاخالی با نسبت ثابت، امتحان شونده باید متن را برای بازسازی و کاربرد توانایی زبانی خود در تکمیل جاخالی‌ها با کلمات مناسب بخواند. این فرآیند در واقع فرآیند درک مطلب و بیان می‌باشد، یعنی امتحان شونده‌گان باید متن را بخوانند و درک کنند و کلمات مناسبی را بیان کنند. به منظور تغییر تکلیف درک مطلب و بیان ناتمام به درک مطلب کامل آزمونهای جاخالی چندگزینه‌ای طراحی شدند.

در این نوع آزمون جاخالی، برای هر جای خالی ۴ گزینه یا بیشتر طرح می‌شود و امتحان شونده‌گان باید مناسبترین گزینه را برای جای خالی انتخاب کنند. آزمون جاخالی چندگزینه‌ای بنابه فرمش دارای مزیت عملی بودن نسبت به آزمون تشریحی می‌باشد. مطالعات حاکی از آن است که آزمونهای جاخالی چندگزینه‌ای اطلاعات مشابه آزمونهای تشریحی را ایجاد می‌کنند. تنها نقص این نوع آزمون حذف بیان امتحان شونده‌گان می‌باشد.

البته حذف تکلیف بیان، دلیل آسانتر بودن این آزمون از آزمون تشریحی می‌باشد. به هر حال این آزمون برای تمام امتحان شونده‌گان آسان می‌باشد، به این ترتیب در قدرت تشخیص آزمون تأثیری ندارد. در برخی موارد، حتی شاید آزمونهای جاخالی چندگزینه‌ای بهتر از آزمونهای تشریحی باشند، یکی از این موارد برگزاری آزمون جاخالی در مقیاس وسیع می‌باشد که در این حالت فرم چندگزینه‌ای عملی‌تر از سایر آزمونها است.

بخش دیگر تحقیق، جهت طرح انواع آزمونهای جاخالی بر فرآیند حذف کلمه متمرکز شد. همانگونه که قبلاً ذکر شد، روش معمول در طرح آزمونهای جاخالی تنها حذف کلمه n ام بود. یعنی n می‌تواند ۵، ۷ و غیره باشد. این روش ثابت گاهی حذف اتفاقی کاذب نامیده می‌شود زیرا میزان n از یک سؤال به سؤال دیگر تغییر نمی‌کند، یعنی میزان n ثابت است. به این علت این روش اتفاقی کاذب نامیده می‌شود که در فرآیند حذف از فرآیند کامل اتفاقی استفاده نمی‌شود. تصور بر این است که زبان به‌طور طبیعی ایجاد می‌شود و رابطه بین کلمات و عقاید بیان شده اتفاقی می‌باشد. اما فرآیند حذف در طرح انواع مختلف آزمونهای جاخالی متنوع می‌باشد.

یکی از انواع آزمونهای جاخالی در رابطه با فرآیند حذف، کاربرد روشهای کاملاً اتفاقی می‌باشد. در این فرآیند تمام کلمات متن نمایانگر یک عدد می‌باشند. (برای مثال ۱ تا ۲۵۰، با فرض این که متن حاوی ۲۵۰ کلمه است) سپس چند عدد به‌طور اتفاقی انتخاب می‌شوند و کلمات مطابق این اعداد حذف می‌شوند. البته ممکن است دو یا چند کلمه مجاور هم حذف شود. اگر چه ممکن است این امر، منجر به تحقیقات جدیدی در زمینه آزمون جاخالی بشود، طراحان آزمون می‌توانند با این شرط که کلمات مجاور هم نباشند کلمات را به‌طور اتفاقی حذف کنند. به علاوه فرآیند اتفاقی کردن در دو مورد به کار می‌رود:

این که جمله اول و آخر دست نخورده باقی بماند یا نه، باز منجر به دو روش دیگر فرآیند جاخالی می‌شود.

نوع دیگر آزمون جاخالی که از تغییر فرآیند حذف ناشی می‌شود بکارگیری فرآیند نمونه‌گیری قشری می‌باشد. در نمونه‌گیری قشری افراد نمونه به تناسب ماهیت جمعیت مورد مطالعه انتخاب می‌شوند. در این روش تعداد کلمات رده‌های دستوری متفاوت تعیین می‌شود. سپس تعداد کلمات حذف به تناسب به هر رده از کلمات اختصاص داده می‌شود. برای مثال تصور کنید که متنی با ۲۴۰ کلمه، دارای ۸۰ کلمه دستوری، ۴۰ فعل، ۴۰ قید، ۵۶ صفت و ۲۴ فعل کمکی باشد. پس (با فرض ۳۰ حذف) ۱۰ کلمه دستوری، ۵ فعل، ۷ قید و ۵ صفت و ۳ فعل کمکی به تناسب حذف خواهد شد. در این روش هر رده از کلمات دارای احتمال مساوی برای قرار گرفتن در فرآیند حذف می‌باشند. انواع آزمونهای جاخالی ذکر شده، دارای شرایط لازم آزمون به‌عنوان ابزار سنجش می‌باشند. از آنجائیکه تئوری‌های مربوط به فرآیند جاخالی بر طبیعی بودن زبان آزمون تأکید دارند، تمام گونه‌های ذکر شده، پدیده طبیعی بودن را با کاربرد روشهای اتفاقی کردن حذف کلمات، رعایت می‌کنند. روشهای حذف با نسبت ثابت، حذف اتفاقی و حذف تصادفی قشری دارای معیار طبیعی بودن می‌باشند، از این رو شرایط نظری فرآیند جاخالی را نیز در بردارند. اما آزمونهای دیگری نیز وجود دارند که دارای معیار طبیعی بودن نمی‌باشند. این نوع آزمونها معمولاً تکالیف جاخالی نامیده می‌شوند نه آزمونهای جاخالی، تکالیف جاخالی را می‌توان با حذف هر کلمه، به هر نسبتی در متنی با هر اندازه طرح کرد.

یکی از انواع تکالیف جاخالی با حذف رده از پیش تعیین شده کلمات طرح می‌شود. برای مثال می‌توان تمام حروف اضافه، حروف تعریف، افعال، صفات و گروههای دیگر کلمات را حذف کرد. یعنی می‌توان یک تکلیف جاخالی را با گروه کلمات متفاوت، یا ترکیبی از دو یا چند گروه طرح کرد. برای مثال، می‌توان یک تکلیف جاخالی طرح کرد که فقط حروف تعریف و حروف اضافه حذف شده باشند.

نوع دیگر تکلیف جاخالی با تغییر تعداد حذف‌ها به‌وجود می‌آید که در نتیجه متنی با طول مختلف ایجاد می‌شوند. برای مثال ممکن است تکلیف جاخالی فقط پنج حذف داشته باشد. به این معنا که طول متن حدود چهار کلمه خواهد بود. حتی شاید متن به جملات واحدی تجزیه شود و هر جمله دارای یک یا دو حذف باشد.

باید به این نکته اشاره کرد که نباید تکالیف جاخالی را به عنوان ابزار سنجش بکار برد. یعنی، تصمیم‌گیری براساس نمرات به دست آمده از چنین تکالیف مناسب نمی‌باشد. تکالیف جاخالی دارای شرایط لازم آزمون می‌باشند، زیرا فرضیه طبیعی بودن زبان را نقص می‌کنند. به عبارت دیگر غیرممکن است که متنی بدون حروف تعریف یا اضافه وجود داشته باشد مگر اینکه بطور تصادفی حذف شده باشد. اگر چه تکالیف جاخالی به عنوان ابزار امتحان به کار نمی‌روند، و نباید هم بروند ولی می‌توان به عنوان فعالیت‌هایی به منظور اهداف آموزشی در کلاس مورد استفاده قرار گیرند. تکلیف پر کردن جاهای خالی، توانایی ذهنی فراگیران را فعال می‌کند. به طور کلی فرآیند جاخالی به عنوان آزمون یا تکلیف، روشی متنوع می‌باشد که برای اهداف آموزشی و ارزیابی به کار می‌رود. فرآیند جاخالی می‌تواند به عنوان فعالیتی معنی‌دار و برای تفهیم مسأله در کلاس مورد استفاده قرار گیرد. به عنوان ابزار ارزیابی، این روش می‌تواند به صورت سنجش تشخیصی برای تعیین نکات ضعف و قوت دانش‌آموزان در زمینه خاصی از زبان عمل کند، همچنین این روش به عنوان آزمون مهارت نیز بکار می‌رود. البته آزمون جاخالی بطور اولیه جهت سنجش مهارت عمومی زبان به صورت کلی مورد استفاده قرار گرفته است. بهر حال، نقش تشخیص این روش به همان اندازه با ارزش گزارش شده است.

۴-۱۵ نمره‌دهی آزمون جاخالی

هنگامی که آزمون جاخالی، بدون توجه به فرمش (یعنی اینکه یا نسبت ثابت، تصادفی یا به گونه‌ای دیگر باشد) طرح می‌شود، بایستی فرآیند نمره‌دهی مشخص شود. در آزمون‌های چند گزینه‌ای هر جای خالی یک سؤال است و یک نمره دارد. به همین ترتیب هر جاخالی به عنوان یک پرسش در نظر گرفته می‌شود و دارای یک نمره است. اگرچه این فرآیند به نظر آسان می‌آید، اما در عمل اینطور نیست. برخلاف آزمون‌های چندگزینه‌ای که گزینه صحیح تشخیص داده می‌شود. در آزمون جای خالی کلمه صحیح باید ارائه شود. بنابراین، امتحان‌کنندگان باید نوع پاسخ صحیح را مشخص کنند. نوع پاسخ، روش‌های مختلف نمره‌دهی را به وجود می‌آورد. انواع متداول نمره‌دهی شامل روش عین کلمات، روش کلمات قابل قبول و روش پاسخ توزینی می‌باشد که هر یک به طور خلاصه در زیر مورد بحث و بررسی قرار گرفته است.

۱-۴-۱۵ روش عین کلمات

عملکرد امتحان‌شوندگان در آزمون جاخالی نشان دهنده توانایی آنها در پر کردن جاهای خالی می‌باشد. هر چه کلمات نوشته شده توسط امتحان‌شونده به کلماتی که نویسنده آزمون در ابتدا به کار برده، نزدیکتر باشد، تطابق بین امتحان‌شونده و نویسنده بیشتر خواهد بود. از این رو، این فرآیند به عنوان روش عین کلمات معروف است که در آن کلمات حذف شده اصلی پاسخ صحیح می‌باشند و اگر امتحان‌شونده آنها را بنویسد نمره می‌گیرد. این آسانترین شیوه نمره‌دهی و عینی‌ترین روش می‌باشد. اما، یافتن

کلمات اصلی حذف شده از متن برای امتحان شونده دشوار است. این امر نه تنها کار را مشکل می‌کند، بلکه تنوع پاسخ‌های ممکن را نیز محدود می‌کند. به علاوه امتحان شونده را از فعال کردن فرآیند خلاق زبانی‌اش در ارائه کلمات مناسب بازمی‌دارد. مشکلات ذکر شده محققین را به‌طرح شیوه دیگر نمره‌دهی که روش کلمات قابل قبول نامیده می‌شود سوق داد.

۲-۴-۱۵ روش کلمات قابل قبول

در اکثر موارد، یک جای خالی را می‌توان با کلمات مترادف کلمه اصلی حذف شده پر کرد. اگر چه برخی معتقدند که کلمه مترادف نمی‌تواند جایگزین دقیقی برای کلمه اصلی باشد. اما به‌نظر منطقی می‌رسد به کلمه‌های مترادفی که جریان طبیعی متن نوشته شده را تکمیل می‌کنند نمره داده شود. یعنی شاید بیش از یک کلمه، برای جای خالی کاملاً مناسب باشد. در نظر گرفتن چنین پاسخ‌هایی به‌عنوان پاسخ صحیح شیوه نمره‌دهی روش کلمات قابل قبول نامیده می‌شود. به‌عبارت دیگر در روش نمره‌دهی کلمات قابل قبول هر پاسخی که به‌نظر گویشور بومی زبان قابل قبول باشد، پاسخ صحیح می‌باشد.

البته، در این روش امتحان شونده حق انتخاب متعدد برای پر کردن جای خالی دارد. اما فرآیند نمره‌دهی مشکل‌تر از روش عین کلمات می‌باشد. یعنی معیار قابل قبول بودن همانند روش عین کلمات، عینی نمی‌باشد. در واقع تعیین کلمه قابل قبول برای بافت خاصی که در آزمون به‌کار رفته است کار آسانی نمی‌باشد. بنابراین بهترین روش، پیش آزمون تست جاخالی در مورد گروهی از گویشگران بومی زبان می‌باشد. پاسخ‌هایی که افراد بومی زبان به سؤال می‌دهند جواب‌های قابل قبولی خواهند بود که قبل از انجام آزمون در مورد افراد غیربومی زبان به‌وجود می‌آید. البته هر پاسخ داده شده توسط گویشگران بومی به‌طور قطعی قابل قبول نمی‌باشد. بنابراین، برای جدا کردن پاسخ صحیح از پاسخ غلط قدرت تشخیص لازم است. به‌همین دلیل است که روش عین کلمات به طرف ذهنی بودن سوق داده می‌شود. علاوه بر ذهنی بودن این روش، تعدادی گویشگر بومی زبان تحصیل کرده برای ارائه کلمات قابل قبول موردنیاز می‌باشد. در موقعیتهای تدریس زبان خارجی که دستیابی به فرد بومی آسان و ممکن نیست، تصمیم‌گیری در مورد میزان قابل قبول بودن کار بسیار مشکلی است. بنابراین روش عین کلمات برای موقعیتهای تدریس زبان دوم مناسبتر از زبان خارجی خواهد بود.

۳-۴-۱۵ روش پاسخ توزینی

نمره دهی به روش پاسخ توزینی از روش عین کلمات گرفته شده است که در آن بعضی از پاسخ‌ها مناسبتر از بقیه می‌باشند. برای مثال هنگامی که کلمه‌ای قواعد جزئی ضوابط بافتی را به‌هم می‌زند، دارای ارزش بیشتری نسبت به کلمه‌ای است که قواعد اصلی ضوابط بافتی را مختل می‌کند. بنابراین

کلمات مختلف دارای نمره‌های متفاوتی هستند. در واقع روشهای توزینی زیادی به وجود آمده و به کار می‌روند.

معروفترین نوع این شیوه را دارنل (۱۹۶۸) با نام نمره جاخالی ارائه داد. نمره جاخالی از تکرر نسبی پاسخها برای سنجش آنها از طریق محاسبات پیچیده ریاضی استفاده می‌کند. اولر (۱۹۷۹) با اشاره به پرسش ویژه جاخالی در بافت خاصی مقیاس کاربردی آسانتری را ارائه داد. به این معنا که یک مقیاس برای تمام آزمونهای جاخالی مناسب نخواهد بود. اصول مقیاس باید در یک آزمون جاخالی به کار رود و سپس میزان قابل قبول بودن و ارزش پاسخها برای این آزمون جاخالی ویژه تعیین شود.

در اینجا باید نکته مهمی در مورد شیوه‌های متفاوت نمره‌دهی ذکر شود. مطالعات متعددی همواره و به طور قطع ثابت کرده‌اند که همبستگی‌های بسیار بالایی (بین ۰/۸۰ تا ۰/۹۵) در میان شیوه‌های متفاوت نمره‌دهی وجود دارد. به نظر می‌رسد، روش نمره‌دهی از متغیرهای مربوط به ماهیت آزمون و فراگیران مستقل می‌باشد. بنابراین در هر شرایطی مسئله عملی بودن به عنوان معیار اصلی قرار می‌گیرد. یعنی برای آزمون زبان انگلیسی به عنوان زبان خارجی روش عین کلمات و برای موقعیت زبان انگلیسی به عنوان زبان دوم روش کلمات قابل قبول مناسبتر خواهد بود. نمره‌دهی پاسخ توزینی به دلیل خسته کننده بودن فرآیند غیر عملی به نظر می‌رسد.

۵-۱۵ آزمونهای املاء

چنانکه قبلاً گفته شد، تکالیف جاخالی و املاء، آزمونهای کاملاً تلفیقی به حساب می‌آیند یعنی در این آزمونها زبان به اجزا و بخشها یا عناصر مجزایی تجزیه نمی‌شود بلکه زبان به صورت کلی ارزیابی می‌شود. اولر (۱۹۷۹) آزمونهای تلفیقی را از آزمونهای اصطلاحاً عملی متمایز کرد. طبق نظر اولر، یک آزمون عملی دارای دو معیار می‌باشد. ابتدا این آزمون به فرآیند توالی موقت عناصر زبانی که از روابط عادی معنی دار چنین عناصری در متنی که تشکیل یافته‌اند، نیاز دارد، یعنی زبان آزمون باید نمونه‌ای از کلام طبیعی باشد نه ساختگی، دوماً آزمون عملی، فرد را ملزم می‌کند تا از طریق ترسیم عملی توالی عناصر را به بافت فرازبانی ارتباط دهد، یعنی زبان آزمون باید مطابق واقعیات جهان بیرون باشد. به این معنا که باید ترسیم یا تطابقی میان عناصر زبانی و بافت فرازبانی وجود داشته باشد.

چنانکه گفته می‌شود، املاء به عنوان ابزار آزمون باید دارای این معیارها باشد و در این صورت آزمون عملی به حساب می‌آید اما این نوع آزمون تاریخچه‌ای طولانی دارد. در تاریخ آزمون زبان به املاء توجه زیادی شده است. در برخی موارد این نوع آزمون مورد انتقاد قرار گرفته در حالی که در موارد دیگر به عنوان ابزاری معتبر برای ارزیابی مهارت زبانی مورد توجه قرار گرفته است.

این انتقادات از جانب ممتحنین قدیمی بوده است. برای مثال لادو (۱۹۶۱) به دلیل جزئی بودن عناصر مورد ارزیابی با املاء مخالفت کرده است. او مدعی بود که چون کلمات و ترتیب آنها مشخص هستند، این روش واژگان و ترتیب کلمات را امتحان نمی‌کند. به همین ترتیب ولت (۱۹۷۶) بر این باور

بود که املاء تفاوت‌های واجی را مورد ارزیابی قرار می‌دهد. هریس (۱۹۶۹) بیشترین مخالفت را در مورد املاء کرده است و مدعی بود که املاء به‌عنوان ابزاری برای آزمون غیردقیق و غیراقتصادی می‌باشد. برعکس انتقادات فوق‌الذکر تحقیق جدیدی که توسط اولر و طرفداران وی انجام گرفته است نشان می‌دهد که املاء یکی از معتبرترین و پایاترین روش‌های ارزیابی مهارت زبانی می‌باشد. کوهن (۱۹۸۰) بر این باور بود که املاء امتحان بافت‌دار هجی کلمات می‌باشد زیرا از قدیم برای امتحان توانایی هجی کلمات به‌کار می‌رود. اما کاربرد جدید املاء جدا از روش قدیمی‌اش می‌باشد. زیرا املاء به‌نحوی که امروزه به‌کار می‌رود متن را کلمه به کلمه ارائه نمی‌کند بلکه بخش به بخش ارائه می‌دهد. روش‌های برگزاری آزمون تفسیر نمرات و تأمین عملی بودن آزمون منجر به پیدایش آزمون به اصطلاح املاء استاندارد همراه با انواع روش‌های دیگر همچون املائی ناقص، تقلید فراخوانی، املاء - انشاء و املاء همراه با صدا شده است.

۱-۵-۱۵ املائی استاندارد

متداولترین نوع املاء، املائی استاندارد می‌باشد. لازم به‌ذکر است که کلمه استاندارد نسبتاً آزاد به‌کار رفته است. شاید بهتر باشد این روش را املائی قراردادی نامید. زیرا اصطلاح استاندارد به‌معیار تعیین شده عملی اشاره دارد. بنابراین در مبحث مربوط به املاء خواننده باید کاربرد کلمه «استاندارد» را به‌صورت اصطلاحی معمولی در نظر بگیرد نه اصطلاحی تخصصی.

املائی استاندارد معمولاً به‌عنوان متنی با طول و سطح دشواری مناسب با نحوه برگزاری خاصی تعریف می‌شود. اکثر محققان در رابطه با طول متن اتفاق‌نظر دارند که متن باید دارای بیش از ۱۰۰ کلمه باشد تا امکان ارزیابی عملکرد امتحان شونده در کلام پیوسته وجود داشته باشد. بنابراین، طول متن چیزی بین ۱۵۰-۱۰۰ کلمه می‌باشد. در مورد سطح دشواری متن همانند روش جاخالی فرمولهای قابلیت خواندن مفید خواهد بود. محققان بر این باورند که متن باید کمی از سطح مهارت دانش‌آموزان پایین‌تر باشد تا مشکل درک مطلب ناشی از خطای ذهنی یا پیچیدگیهای زبانی رفع شود و بالاخره، متن انتخاب شده باید طبق روش‌های مقرر شده خاصی ارائه شود.

معمولاً املاء خواننده می‌شود و از داوطلبین خواسته می‌شود آنچه را که می‌شنوند، بنویسند. با توجه به این روش دو نکته مهم وجود دارد. ابتدا یکنواختی خواننده و شیوه خواندن. برای اجتناب از اشتباهات فردی و دستیابی به یکنواختی خواننده روش پیشنهادی ضبط متنی است که توسط فراگیر بومی زبان خوانده می‌شود. این کار بسیاری از مشکلات مربوط به تفاوت تکیه، تلفظ، مکث، فصاحت، تفاوت‌های زیر زنجیری بین خوانندگان املاء را کاهش می‌دهد. همچنین اگر قرار باشد، یک خواننده متن را دوبار یا برای دو گروه مختلف بخواند فرآیند خواندن یکنواخت خواهد بود.

دوماً، متن املاء باید سه بار خوانده شود. هر خواندن هدف خاصی را تأمین می‌کند. هنگامی که متن برای بار اول خوانده می‌شود، امتحان شوندگان چیزی نمی‌نویسند، بار دوم مهمترین بخش امتحان

می‌باشد. زیرا در این مرحله آنها باید چیزی را که می‌شنوند بنویسند. در بار دوم، باید متن با سرعت معمولی، مکشهای مناسب و علائم نقطه گذاری لازم خوانده شود. چون مکشها در متنهای طبیعی وجود دارند، طول این مکشها بنابه تعداد کلمات موجود در هر بخش تغییر می‌کند. بنابراین قبلاً تعیین طول هر مکث در رابطه با تعداد کلمات کار مهمی می‌باشد. توصیه می‌شود که محدوده بخشها و زمان لازم برای آنها برای خواننده متن مشخص شود. بالاخره مرحله سوم به همان شیوه خواندن اول انجام می‌گیرد. هدف از خواندن برای بار سوم این است که به دانش‌آموزان اجازه دهد تا کارشان را کنترل کنند.

نکته مهمی که در رابطه با خواندن املاء لازم به ذکر است این است که در هر سه مرحله خواندن باید متن با سرعت معمولی خوانده شود و هیچ کد خوانی عمدی و یا غیرطبیعی جایز نمی‌باشد. زیرا هدف املاء ارزیابی توانایی فرد در حالت طبیعی است.

بنابراین به‌دلیل ماهیت کلام طبیعی، بخشهای داده شده در یک فاصله زمانی از یک کلمه تا جمله کامل متغیر می‌باشند. البته این املاء از املاهای کلمه به کلمه‌ای که معمولاً در اکثر مؤسسات آموزشی انجام می‌گیرد، کاملاً متفاوت است. در حقیقت با مختل کردن معیار طبیعی بودن، املاهایی که با فرم و سرعت طبیعی انجام می‌گیرند، بیشتر آزمون هجی کلمات خواهند بود تا املاء.

۲-۵-۱۵ املاهای ناقص

در املاهای ناقص، متنی با چند حذف در اختیار امتحان شوندگان قرار می‌گیرد، اما متن به‌طور کامل خوانده می‌شود. امتحان شوندگان باید با شنیدن متن، بخشهای افتاده را پر کنند. در واقع املاهای ناقص فعالیتی است که در حد فاصل تکالیف جاخالی و املاء قرار می‌گیرد.

از آنجائیکه متن برای امتحان شونده خوانده می‌شود، این کار به املاء شباهت دارد اما از این نظر که متن ناقص در اختیار امتحان شوندگان قرار می‌گیرد، با آن متفاوت است. همچنین این نوع املاء از جهت پر کردن جاهای خالی به آزمون جاخالی شبیه می‌باشد اما چون بخشهای حذف شده طی خواندن متن به امتحان شونده داده می‌شود از آزمون جاخالی متفاوت است.

املاء ناقص با افزایش طول بخشهای حذف شده، مشکل‌تر می‌شود. یعنی هر چه بخشهای حذف شده بلندتر باشد، کار دشوارتر خواهد شد. همچنین سرعتی که متن خوانده می‌شود بر میزان دشواری آزمون تأثیر دارد. یعنی هر چه متن سریعتر خوانده شود کار مشکل‌تر خواهد بود.

۳-۵-۱۵ تقلید فراخوانی

در تقلید فراخوانی، امتحان شوندگان باید چیزی را که می‌شنوند، تقلید کرده و تکرار کنند. دلیل فراخوانی بودن این کار، این است که امتحان شوندگان سعی می‌کنند اطلاعاتی را کسب کنند اگر چه

برخی این کار را کاری سطحی و ماشینی می‌پندارند اما این عمل حافظه کوتاه مدت امتحان شوندگان را که نیاز به پردازش عمیق دارد فعال می‌کند.

این کار به این دلیل در زمره فعالیت‌های املایی قرار می‌گیرد که امتحان شوندگان باید چیزی را که می‌شنوند بیان کنند. البته نه به شکل نوشتاری مثل املای استاندارد بلکه به صورت شفاهی. این شیوه در مورد کودکان و بزرگسالان نوآموز بسیار مؤثر است. اما عملاً در آزمونهای آموزشی هم قابل استفاده می‌باشد.

۴-۵-۱۵ املاء - انشاء

چنانکه از خود کلمه برمی‌آید، املاء - انشاء از ترکیب املاء و انشاء تشکیل یافته است. در املاء - انشاء مطالب به صورت شفاهی با سرعت معمولی به امتحان شوندگان ارائه می‌شود و آنها باید چیزی را که شنیده‌اند، بنویسند. از این نظر که مطالب به‌طور شفاهی داده می‌شود این کار به املاء شبیه است از آنجائیکه مطلب به‌صورت کلی ارائه می‌شوند نه به‌صورت قطعه‌قطعه کاری متفاوت از املاء می‌باشد. وجه تشابه این کار با انشاء در این است که امتحان‌شوندگان باید متنی را که می‌شنوند به صورت انشاء بنویسند و تفاوتش با انشاء در این است که مطالب در اختیار امتحان‌شوندگان قرار می‌گیرد.

همچنین املاء - انشاء تفاوت‌هایی با املای ناقص دارد. از این نظر که در اولی امتحان‌شونده باید بعد از گوش کردن متن کاملی بنویسد در حالی‌که در دومی باید با گوش کردن بخش‌های خاصی را بنویسد. در واقع املاء - انشاء کار پیچیده‌ای است که شامل گوش کردن، فهمیدن، پردازش کردن، بازسازی کردن، سازماندهی و در نهایت نوشتن مطالب می‌باشد. این شیوه برای دانش‌آموزان سطوح متوسط و پیشرفته مناسب می‌باشد.

۴-۵-۱۵ املا با سروصدا

این نوع املاء با افزودن صدای خفیف به محتوای مطالبی که خوانده می‌شود یا از نوار پخش می‌شوند انجام می‌گیرد. صدای خفیف به طیف صدای اتفاقی گفته می‌شود که به پارازیت رادیو شبیه است و وضوح مطالب املاء را برهم می‌زند. امتحان‌شوندگان زبان را در شرایطی مشابه واقعیت دریافت می‌کنند یعنی شرایط صوتی مطلوب همواره در طبیعت وجود ندارد. برای مثال، شرایطی مثل گوش کردن به مکالمه در اتاقی که رادیو و تلویزیون روشن می‌باشد یا گوش کردن به صحبت فردی در جای شلوغ مثالهایی از تداخل صدا با کارهای پردازش می‌باشند. شدت صدا در مطالب معمولاً به عنوان صدا به نسبت سیگنال گفته می‌شود. هر چه این نسبت بیشتر باشد اختلال بیشتر و در نتیجه مهارت زبانی بالاتری برای درک مطلب لازم خواهد بود.

۶-۱۵ نمره‌دهی آزمونهای املاء

نمره‌دهی آزمونهای املاء یکی از بحث‌انگیزترین موضوعات در چند دهه گذشته بوده است، زیرا تغییری جزئی در فرآیند برگزاری، کار را به‌شدت آسان یا مشکل می‌سازد. علاوه بر عوامل عادی مؤثر در دشواری کار، یعنی واژگان، پیچیدگی نحوی و تغییرات سبکی در متن، عوامل دیگری وجود دارند که در دشواری یا آسانی کار مؤثر هستند.

عامل اصلی که کم و بیش در دشواری املاء تأثیر دارد، سرعت کلی ارائه مطالب می‌باشد. هر چه متنی سریع‌تر ارائه شود کار مشکل‌تر خواهد شد. عامل دیگری که در میزان دشواری کار تأثیر دارد طول مطالب ارائه شده بین مکثها می‌باشد. هر چه زنجیره کلمات بین دو مکث طولانی‌تر باشد بار بیشتر بر حافظه کوتاه مدت داوطلب تحمیل می‌شود و نهایتاً کار مشکل‌تر می‌شود. آخرین و مهمترین عامل تعداد دفعاتی است که مطالب خوانده می‌شود و هر چه تعداد دفعاتی که متن خوانده می‌شود، بیشتر باشد، متن بیشتر درک می‌شود و به این ترتیب کار آسان‌تر خواهد شد.

این عوامل می‌توانند املاء را از کاری آسان به کار بس مشکل تبدیل کنند. اگر چه برخی از این تغییرات با در نظر گرفتن روشهای ذکر شده در املاي استاندارد کاهش می‌یابند، اما به‌طور قطعی قابل کنترل نمی‌باشند. بنابراین طرح معیاری قطعی برای برگزاری امتحان و نهایتاً نمره‌دهی املاء امری بسیار مشکل است.

روشهای نمره‌دهی بسیاری ارائه شده‌اند. متداولترین روش نمره‌دهی روش کلمه به کلمه می‌باشد. یعنی هر کلمه در املاء یک سؤال به حساب می‌آید و اگر کلمه‌ای صحیح نوشته شود یک نمره برای محسوب می‌شود. یکی از ویژگیهای مهم این روش نمره‌دهی در این است که اشتباهات املائی تا جایی که معنی کلمات حفظ شود غلط به حساب نمی‌آید. برای مثال نوشتن کلمه "beautiful" به صورت "beuteful" که سه اشتباه املائی وجود دارد، غلط محسوب نمی‌شود، و نمره‌ای از امتحان شونده کم نمی‌گردد. اما نوشتن "ship" به جای "sheep" یا "to" بجای "two" و یا "right" به عوض "write" حتی با یک اشتباه املائی، غلط حساب می‌شود زیرا معنی کلمه تغییر یافته است. این فرآیند معمولاً در نمره‌دهی آزمونهای املاء به کار می‌رود.

همچنین گونه‌های متعدد نمره‌دهی دیگری نیز وجود دارند یکی از این روشها با کمی تفاوت مانند روش قبلی می‌باشد. یعنی برای اشتباهات املائی نمره کم می‌شود. این روش مشابه روشهایی است که معمولاً در نمره‌دهی املاهای سنتی به کار می‌رفت. در نوع دیگر تعداد حروف کلمه به عنوان واحد نمره‌دهی در نظر گرفته می‌شود. اصولاً این روش، روشی پایاتر می‌باشد. زیرا طول کلمه در نظر گرفته می‌شود. در واقع تحقیقات اخیر این فرضیه را اثبات می‌کند که هر چه واحد نمره‌دهی عینی‌تر باشد، پایایی و اعتبار املاء افزایش می‌یابد. آخرین روش هر بخش داده شده به امتحان شونده را به عنوان یک

سؤال به حساب می‌آورد. اگر چه این روش به دلیل نزدیکی به معیار طبیعی بودن، قابل توجه‌تر از روشهای دیگر می‌باشد. اما نسبت به سایر روشهای دیگر ذهنی‌تر و وقت‌گیرتر می‌باشد. بنابراین، پیشنهاد قطعی هر یک از شیوه‌های نمره‌دهی فوق کار مشکلی می‌باشد. بررسی‌های بیشتری برای روشن شدن موضوع باید صورت گیرد اما تا دستیابی به نتیجه‌ای قطعی کاربرد شیوه کلمه به کلمه پیشنهاد می‌شود.

فعاليتها

۱- روشهای زیر همچنین به عنوان شیوه‌های دیگر آزمون جاخالی پیشنهاد شده‌اند. نظر شما در مورد هر یک از سؤالات چیست؟ مزایا و معایب هر کدام را ذکر کنید.

الف) برای مثال، روشی شبیه روش جاخالی وجود دارد. در این روش که به نام آزمون C شناخته شده است، اولین نیمه هر دومین کلمه حذف می‌شود. اگر تعداد حروف کلمه فرد باشد، نصف کلمه با اضافه یک حذف می‌شود. کلمات تک حرفی مثل، "I" و "a" به حساب نمی‌آیند و شمارش از کلمه بعدی شروع می‌شود. امتحان شوندگان حروف حذف شده را کامل می‌کنند. فقط کلماتی که کاملاً تکمیل شده باشد، درست به حساب می‌آیند. نمونه‌ای از آزمون C در اینجا آورده شده است.

One upon a time, there was a little. Girl who lived with her mother who was a widow. They were so p that o day in th had not Left t Eat. T..... little gi want to Into woods t..... play. S.... was s.... hungry th..... she be to. C..... An , o woman ca..... up t.... her. We are y Canying, m..... child. S.... asked. Be c.... I am hungry. Said the little girl, then you shall be hungry no more, said the old woman.

ب) مکالمه کوتاهی با یک جای خالی کلمه یا عباراتی در اختیار امتحان شوندگان قرار می‌گیرد. آنها باید از میان چهار گزینه داده شده. گزینه مناسب را انتخاب کنند. برای مثال:

Was that T.V program interesting?
I thought it was boring.
What did your friends think of it.
They were, too.

a) tired

b) boored

c) interested

d) enter tained

۲- یک آزمون جاخالی با استفاده از روش حذف سیستماتیک طرح کرده، در مورد گروهی از دانش‌آموزان برگزار کنید. سپس شاخصهای دشواری و افتراق هر جای خالی را تعیین کنید. چگونه می‌توان این آزمون را اصلاح کرد.

۳- دو روش آماده سازی آزمون جاخالی چند گزینه‌ای قبلاً ذکر شد. آیا روشهای دیگری نیز وجود دارند؟ کدام روش را ترجیح می‌دهید؟ چرا؟

۴- املای استاندارد را با املا - انشاء مقایسه کنید. کدام یک برای دانش آموزان سطوح پیشرفته مناسب‌تر می‌باشد؟

۵- چه رابطه‌ای بین آزمونهای جاخالی و املاء از لحاظ فرآیندهای روانشناسی وجود دارد؟

آزمون‌های کارکردی

۱-۱۶ مقدمه

همانطور که قبلاً اشاره شد در زمینه آزمون زبان، دو نظریه مختلف وجود دارد. اولین نظریه که تقریباً بیش از نیم قرن مطرح بوده است، روش مجزا (شکل) نامیده می‌شود. دومین نظریه که از کمبودهای جدی روش اول تکوین یافت روش تلفیقی نامیده می‌شود. مطالعه و بررسی تاریخچه این دو روش ظاهراً متفاوت، بنظر ضرورتی نمی‌رسد. اگر چه به منظور بحث و بررسی بیشتر، توضیحات مختصر مفید خواهد بود.»

روش مجزا - شکل (تجربیدی) از این باور بوجود آمد که زبان سیستمی از گروه‌های مجزایی همچون واجها، تکواژها، کلمات، عبارات و جملات می‌باشد. طرفداران این روش معتقدند که آزمودن این زنجیره‌ها، به معلمین این امکان را می‌دهد تا مهارت زبانی امتحان شوندگان را ارزیابی کنند.

پس از چند دهه حاکمیت روش مجزا شکل، این روش مورد انتقاد شدید و غیرمنطقی محققان قرار گرفت. مخالفان آزمون مجزا - شکل بر این باور بودند که، چون نقش اصلی زبان، برقراری ارتباط است بنابراین آزمون‌های زبان باید با توانایی ارتباطی فراگیران نیز مرتبط باشد. آنها همچنین اظهار می‌کردند که تاثیر هر سوال مجزا در کل مهارت زبانی نه تنها ارزشمند بلکه قابل تشخیص هم نیست. بنابه این دلیل و خیلی از دلایل دیگر، گرایش به سوی طرح آزمون‌های تلفیقی، با امید به اینکه چنین آزمون‌هایی توانایی ارتباطی زبان را مورد سنجش قرار دهند، تغییر کرد.

تا زمانیکه نظریه آزمون‌های کاربردی (نسبت به آزمون‌های موجود) بعنوان بهترین روش شناخته شد، این اختلاف نظرها همچنان ادامه داشت. اگر چه بنظر می‌رسد که بین این شیوه (عملی و

یکپارچه) تفاوتی وجود ندارد و اگر هم هست بسیار ناچیز است. در اکثر موارد آزمون‌های عملی زیر مجموعه‌ای از آزمون‌های یکپارچه با ویژگی‌های کاملاً مشخص به حساب می‌آیند.

بدیهی است که ابعاد مختلف آزمون از تاثیر شیوه‌های مختلف تدریس ناشی می‌شود، یعنی تحولات جدید در تدریس زبان، درست یا غلط منجر به پیدایش روش‌های مختلف آزمون شده است. بدنبال این رشته مباحث، آزمون مجزا شکل از شیوه‌های تدریس مجزا مثل روش شنیداری - گفتاری، شکل گرفته است. به همین ترتیب شاید آزمون‌های تلفیقی یا یکپارچه از تاثیر روش‌های تدریس شناختی بوجود آمده باشد. صرفنظر از اینکه کدام روش بکار می‌رود، می‌توان گفت که هیچ یک از این شیوه‌های تدریس و همچنین هیچ یک از روش‌های آزمون سازی بطور کامل به عملکرد ارتباطی فراگیران زبان توجه نکرده است. این نارسایی می‌تواند یکی از دلایل پیدایش روش‌های جدید تدریس باشد. روشی نسبتاً جدید که روش کارکردی - مفهومی نامیده می‌شود در اروپا شکل گرفت و مورد توجه کارشناسان انگلیسی بعنوان زبان دوم قرار گرفت. شیوه کارکردی مفهومی سعی بر آن دارد تا مباحث کارکردهای ارتباطی را تعیین و تدریس کند. این کارکردها مشابه کاربرد واقعی زبان در موقعیت‌های اجتماعی خاص می‌باشند. ویکنز (۱۹۷۶) یکی از طرفداران این روش چنین اظهار می‌کند: «الزام کلی روش مفهومی نسبت به تدریس زبان از این واقعیت نشأت می‌گیرد که چیزی که افراد می‌خواهند از طریق زبان انجام دهند، مهمتر از مهارت در زبان بعنوان یک سیستم غیر کاربردی است.»

پیدایش بُعد جدید در تدریس زبان نیازمند وجود بُعد جدید در زمینه آزمون سازی زبان می‌باشد. اما به دلیل پیچیدگی فرآیندهای کارکردی و تنوع مقوله‌های کاربردی، هیچ گونه پیشرفت واقعی در توسعه آزمون‌های عملکردی صورت نگرفته است. ضرورت وجود چنین آزمون‌هایی قبلاً مطرح شده و برخی راهکارهای سودمند توسط محققان ارائه شده است. با این وجود، هنوز پیشرفت واقعی آزمون‌های کارکردی تحقق نیافته است. بقیه این فصل به توجیه موقعیت و کاربرد آزمون‌های کارکردی در زمینه آزمون مهارت زبانی اختصاص یافته است. بنابراین مشکلات و معایب آزمون‌های موجود مورد بحث و بررسی قرار خواهد گرفت، پس روشهای طرح آزمون‌های کارکردی ارائه خواهد شد.

۲-۱۶ مشکلات آزمون‌های مجزا شکل

اگر چه تعریف دقیقی از نوع سوالات تشکیل دهنده آزمون مجزا شکل ارائه نشده است، اما این نظریه کلی وجود دارد که سوالات آزمون مجزا شکل، بخش خاصی از زبان را در یک زمان مورد ارزیابی قرار می‌دهد.

همچنین میزان مجزا بودن دارای سطوح مختلفی می‌باشد و یک سوال را می‌توان از بیشتر مجزا بودن به کمتر مجزا بودن طبقه‌بندی کرد. تمام آزمون‌های مجزا شکل دارای چنین مراحل طرح پیش آزمون و بازنگری می‌باشند که در پایان به آزمون‌های ظاهر پایا و معتبر منتهی می‌شوند. اما صرفنظر از

پایایی که مفهومی جدا از محتوای آزمون بوده و براساس محاسبات ریاضی می‌باشد، اکثر محققان اعتبار آزمون‌های مجزا شکل را مورد سوال قرار داده‌اند.

وقتی اعتبار آزمونی زیر سوال می‌رود بنابراین در مورد پایایی آن آزمون نیز تردید بوجود می‌آید، بدون وجود اعتبار آزمون تمام معیارهای دیگر نیز از جمله پایایی فاقد ارزش خواهند بود. البته چنین استنتاجی بدین معنا نیست که آزمون‌های مجزا شکل در ارزیابی مهارت زبانی نقشی ندارند. بالعکس با در نظر گرفتن مشکلات مربوط به آزمون‌های تلفیقی، آزمون‌های مجزا شکل در میان آزمون‌هایی که برای اهداف خاص و یا اهداف تشخیصی طراحی شده‌اند، از جایگاه مطلوبی برخوردارند.

..... تاکنون آزمون‌های تلفیقی از جهت توجیه پذیری و پذیرفتگی مناسب‌ترین جایگزین برای آزمون‌های مجزا شکل بوده‌اند رایج‌ترین نوع آزمون‌های تلفیقی آزمون‌های فهم و ادراک (جاهای خالی متن را باید با واژه‌های مناسب پرکرد) و املاء می‌باشند که در فصل قبل مورد بررسی قرار گرفتند. اگر چه این آزمون‌ها مورد تایید محققان قرار گرفته‌اند و شاخص‌های پایا و معتبری از خود نشان داده‌اند، اما این آزمون‌ها دارای برخی نقایص عملی و نظری می‌باشند. در صفحات بعدی آزمون‌های جا خالی و املاء بطور دقیق بررسی شده و نقایص عملی و آماری آنها مورد بحث قرار خواهند گرفت.

۳-۱۶ مشکلات آزمون‌های تلفیقی

این عقیده وجود دارد که املاء یکی از روشهای مناسب در ارزیابی مهارت کلی زبان است. اغلب گفته می‌شود که املاء همچون آزمون‌های درک مطلب شنیداری، پایا و معتبر می‌باشد و همچنین گفته است که آزمون‌های املاء، نتایج دقیقی ارائه می‌دهند و این آزمون‌ها نمونه‌های اتفاقی (بدون نقشه قبلی) از کارهای زبانی بافت‌دار می‌باشند.

در صحت این مطالب هیچ تردیدی وجود ندارد چون این گزارشات براساس تحقیقات تجربی می‌باشند. اما نکته قابل سوال، روش تفسیر این نتایج می‌باشد. چنانچه قبلاً ذکر شد املاي استاندارد، روخوانی متن با سرعت معمولی و مکث‌های مناسب در جاهای لازم می‌باشد تا امتحان شونده بتواند متن را بنویسد. بنظر می‌رسد این آزمون با آزمون گوش کردن و درک مطلب که جملات بطور مستقل خوانده می‌شوند، تفاوت زیادی نداشته باشد.

اما ارتباط جملات با یکدیگر، متن بافت‌داری بوجود می‌آورند که در مقایسه با آزمون‌های مجزا شکل به حالت طبیعی زبان نزدیکتر می‌باشد. با وجود این مزایا آزمون املاء بعنوان ابزار امتحان نقایصی را دربردارد.

..... صرف نظر از مشکلات عملی از جمله زمان نمره‌دهی و تعیین بخش‌های تشکیل‌دهنده یک پرسش، بنظر می‌رسد آزمون‌های املاء مشکلاتی داشته باشند. مشکل اول این است که نمرات آزمون املاء به دو عامل بستگی دارد: توانایی درک کلام گفتاری و توانایی شنیدن مطالب درک شده. بدیهی است که عدم توانایی در نوشتن مطالب با سرعتی که شنیده می‌شوند و درک می‌شوند هیچ‌گونه ارتباط

مستقیمی با مهارت زبانی ندارد. تندنویسی مهارت مکانیکی می‌باشد. در حالیکه درک یک پیام، مهارت شناختی به حساب می‌آید. بنابراین نمرات بدست آمده از املاء تا حد زیادی تحت تأثیر مهارت تندنویسی قرار می‌گیرند. ثانیاً از آنجائیکه هر کلمه در نمره‌دهی از ارزش یکسانی برخوردار است، و کلمات بلند زمان بیشتری را به خود اختصاص می‌دهند، امتحان شونده می‌تواند آنها را ننویسد. بنابراین پاسخ به سوالات املاء (با این فرض که هر کلمه یک سوال است) به عامل بالقوه نامربوط طول بستگی دارد.

این مشکلات در اعتبار املاء که میزان ارزیابی آزمون از مطالب موردنظر را تعیین می‌کند تأثیر خواهند داشت. اینکه املاء توانایی گوش کردن و درک مطلب یا سرعت نوشتن مطالب درک شده و یا طول کلمات یا ترکیبی از هر سه را مورد ارزیابی قرار می‌دهد یا نه، مشخص نمی‌باشد. اگر پاسخ این سوال آخرین احتمال ممکن باشد، پس نمرات املاء چگونه تفسیر می‌شوند؟ منظور از این نمرات چیست؟ و تا چه اندازه در معلومات ما از مهارت زبانی امتحان شوندگان سهم خواهند داشت؟

برخلاف املاء، مشکلات مزبور در رابطه با اعتبار آزمون‌های جاخالی، به اندازه مشکلات املاء مهم نمی‌باشند، زیرا آزمون‌های جاخالی نسبت به املاء کنترل شده‌تر می‌باشند. یک آزمون جاخالی دارای پرسش‌های یکسانی در تمام متن می‌باشد که در محاسبه ضریب پایایی و اعتبار بکار می‌روند. اما این سوالات توسط طراحان آزمون تعیین نمی‌شوند و تغییر جزئی در تمام سوال‌ها تأثیر کرده و هدف کلی آزمون را تغییر خواهد داد و مسلم است که هیچ نظریه واضحی در مورد مطالب مورد ارزیابی آزمون جاخالی وجود ندارد و در صورتی که مطالب مورد ارزیابی آزمون معلوم نباشد، پس طرح موضوع اعتبار مناسب خواهد بود و از آنجائیکه اعتبار رابطه نزدیکی با مفهوم پایایی دارد، می‌توان گفت که ضرایب پایایی که برای آزمون جاخالی و املاء بدست آمده‌اند، احتمالاً گمراه کننده می‌باشند. در صورتی که آزمون ناپایا نتواند معتبر باشد، این آزمون‌ها دارای مشکلات جدی خواهند بود که مورد بررسی قرار می‌گیرند.

۱-۳-۱۶ مشکلات مربوط به پایایی

ویژگیهای خاص وجود دارند که هر یک از ابزارهای ارزیابی باید دارای این ویژگیها باشند دو ویژگی مهم آزمون. پایایی و اعتبار آن هستند.

پایایی آزمون‌های جاخالی و املاء معمولاً با استفاده از معیارهای ثبات درونی تعیین می‌شوند. از آنجائیکه این آزمون‌ها فرضیه عدم وابستگی سوالات براساس ضرایب ثبات درونی را نقض کردند، تفسیر قبلی پایایی این آزمون‌ها مورد سوال قرار می‌گیرند.

منظور از عدم وابستگی سوالات این است که عملکرد در یک سوال نباید در عملکرد سوال دیگر تأثیر داشته باشد. اما سوالات املاء و آزمون جاخالی از نظر بافتی به یکدیگر وابسته‌اند یعنی اگر امتحان

شونده نتواند به سوال مهمی پاسخ دهد او سوالات دیگر را هم از دست خواهد داد، زیرا پاسخ به یک سوال شاید تا حدی به پاسخگویی سوال دیگر نیز مربوط باشد.

نظریه تاثیر بافت در عملکرد امتحان شونده توسط زیکو (۱۹۷۸) مطرح شده است.

او آزمون‌های جاخالی را در سه شکل برگزار کرد: آزمون کامل، آزمونی که به ۵ بخش مستقل تقسیم شده و آزمونی که به ۲۰ بخش مستقل تقسیم شده است. وی در عملکرد یک گروه در انواع آزمون‌ها به تفاوت‌های مهمی پی برد. افرادی که در آزمون کامل آزمایش شدند عملکرد بهتری داشتند، زیرا آنها در پاسخ به سوالات، از متن استفاده می‌کردند. برخی محققان دیگر چنین اظهار کردند که عواملی چون چگونگی اولین سوال و محتوای متن بطور بالقوه در عملکرد امتحان شوندگان در آزمون جاخالی تاثیر دارد.

چالش دیگری که در رابطه با آزمون‌های جاخالی و املاء وجود دارد این است که این نوع آزمون‌ها توانایی ارتباط زبانی امتحان شوندگان را ارزیابی نمی‌کنند. این آزمون‌ها به‌طرز موثری عملکرد اجتماعی - فرهنگی، اجتماعی - زبانی و ارتباطی فراگیران را ارزیابی نمی‌کنند. این نوع آزمون‌ها همچون آزمون‌های مجزا شکل در رابطه با ارتباط بین دو یا چند مصاحب، مناسب نمی‌باشند.

بررسی مشکلات بالقوه اساسی در مورد آزمون‌های تلفیقی و نارسایی‌های آزمون‌های مجزا شکل که قبلاً مورد بحث و بررسی قرار گرفتند، نشان می‌دهد که هر دو نوع آزمون برای ارزیابی مهارت زبانی نامناسب می‌باشند. اما به دلیل تنوع مهارت‌های زبانی تحت امتحان و اختلاف نظر در مورد ماهیت آزمون‌های زبان این نظریه بطور معمول پذیرفته شده است که آزمون زبان متعادل باید شامل هر دو بخش تلفیقی و مجزا شکل باشد. اما نمی‌توان پذیرفت که یک آزمون متعادل هدف اصلی آزمون‌های زبانی باشد و باید دنبال معیارهای پارامتر و معتبرتری بود.

یک روش ممکن، طرح آزمون‌هایی می‌باشد که در آن مشکلات مربوط به آزمون‌های مجزا شکل، تلفیقی بطور سیستماتیک به حداقل رسیده باشد، یعنی آزمون‌هایی که ویژگی‌های سوال مستقل را از تئوری مجزا شکل و کاربرد بافتدار زبان را از تئوری تلفیقی انتخاب کنند. در صفحات بعد، روش‌های لازم برای ساخت چنین آزمون‌های مورد بررسی قرار گرفته است.

۴-۱۶- فرآیندهای طرح آزمون‌های کارکردی

دلیل انتخاب اصطلاح کارکردی این است که این نوع آزمون‌ها باید طبق اصول و روش تدریس نقشی - مفهومی باشند. چنانکه قبلاً بحث شد دلیل عمده پیدایش شیوه نقشی - مفهومی، نارسایی‌های متعدد شیوه‌های موجود در رابطه با کارکردهای ارتباطی زبان در موقعیت اجتماعی مناسب بود.

نظریه موجود در شیوه نقشی - مفهومی از نیازهای ارتباطی پیش‌بینی شده فراگیران و تلاشهایی که برای تعیین و تدریس فعالیت‌های زبانی که فراگیران شاید در موقعیت‌های واقعی زندگی با آنها

مواجه شوند، ناشی می‌شود به عبارت دیگر شیوه نقشی - مفهومی کانون توجه آموزشی را از زبانشناسی به نیازهای ارتباطی فراگیران تغییر داد.

بدیهی است که شناسایی و طبقه‌بندی کارکردهای زبانی هرگز جامع نمی‌باشند. قابل ذکر است که تعداد نامحدود الگوهای کارکردی مورد آزمون، مشکلات مشابهی چون آزمون‌های مجزا شکل را بوجود می‌آورند. در آزمون‌های کارکردی صرفنظر از تعداد یا نوع قوانین زبانی، هر سوال دارای روش ارتباطی مستقل می‌باشد. بنابراین صرفنظر از این که کارکرد چیست، نیازی به بررسی تاثیر این کارکرد در فرآیند ارتباطی وجود ندارد و این امر در مورد سوالات آزمون مجزا شکل صدق نمی‌کند زیرا این پرسش‌ها دارای قوانین مجزایی می‌باشند که هیچ اشاره‌ای به کاربرد بالفعل و بالقوه آن قوانین در ارتباط واقعی ندارد.

بنابراین آزمونی که طبق شیوه نقشی - مفهومی باشد باید دارای سوالاتی باشد که توانایی امتحان شوندگان را در رابطه با کارکرد ارتباطی زبان ارزیابی کند تا توانایی آنها در کاربرد فرم‌های زبانشناسی. البته این بدان معنا نیست که صحت زبانی یک جمله مورد توجه نیست. بالعکس اگر چه میزان ارزش معیارهای دستوری مشخص نمی‌باشد صحت زبانی به اندازه رفتارهای زبانی دیگر دارای اهمیت می‌باشد.

با توجه به محتوای سوالات کارکردی، اغلب محققان براین باورند که چنین پرسش‌هایی باید برای ارزیابی مهارت‌های تلفیقی طرح شوند تا مهارت‌های مجزا.

یک آزمون کارکردی باید به سوالاتی از این قبیل پاسخ‌گو باشد، منظور از توانایی امتحان شونده برای انجام کارکردی خاص در شرایط اجتماعی مناسب چیست؟ بنابراین در طرح آزمون کارکردی، انتخاب کارکردها از اهمیت بالایی برخوردار خواهد بود. علاوه بر انتخاب کارکردها، مراحل دیگر در طرح آزمون کارکردی، از جمله تعیین عوامل اجتماعی، تعیین معیارهای عملکرد، طرح پرسش‌های آزمون، و پیش آزمون و بازنگری پرسش‌ها باید طی شوند. هر یک از این مراحل بطور مختصر مورد بررسی قرار گرفته‌اند.

۱-۴-۱۶ انتخاب کارکردها

رده‌های متعددی از کارکردها و ریز کارکردهای زبانی با عناوین مختلفی وجود دارند. هر رده از کارکردهای زبانی را از دیدگاه‌های مختلفی بررسی کنیم، اگر چه اکثر آنها در مقوله‌های اصلی دارای وجوه مشترکی می‌باشند.

تاکنون جامع‌ترین فهرست کارکردهای زبانی، فهرستی است که راگ (۱۹۷۶) با کمک تعداد زیادی از محققان زبانشناسی کاربردی ارائه داده است.

البته برخی از اساتید اصول مربوط به این طبقه‌بندی را مورد انتقاد قرار داده و موفقیت این کارکرد را در تدریس زیر سوال برده‌اند و گروهی دیگر بر این باورند که ریز کارکردها فقط در حد فهرست‌های

ساده نمی‌باشند بلکه به روش‌های مختلفی درجه‌بندی شده‌اند. اما این مخالفت‌ها تاثیر مهمی در طرح آزمون ندارد، زیرا آزمون برای ارزیابی رفتارهای زبانی در شرایط ارتباطی صرفنظر از ترتیب مقوله‌ای یا سلسله مراتب کارکردها طراحی می‌شود. طبقه‌بندی کارکردها توسط وان راک به‌منظور اهداف تدریسی صورت گرفته است یعنی این کارکردها و مفهوم‌ها اهداف آغازین روش مفهومی - کارکردی را تشکیل می‌دهند. بنابراین این طبقه‌بندی دلیل مناسبی برای انتخاب کارکردها و گنجاندن آنها در آزمون می‌باشد.

طبق طبقه‌بندی وان راک، شش کارکرد کلی وجود دارد که هر یک چندین ریز کارکرد دارند (از چهار تا بیست و شش). پرداختن به تمام کارکردها و ریز کارکردها از حیطه یک آزمون خارج است. بنابراین تعداد کارکردها باید تا حد قابل کنترلی محدود شوند. البته، مشکل انتخاب کارکردها و ریز کارکردها با طرح معیارهای آزمایشی تا حدی قابل رفع می‌باشد. برای مثال برای این‌که بتوان معیارهایی را وضع کرد باید کارکردها شامل:

۱- فرم‌های زبانی مختلف (برای مثال: روش‌های نحوی مختلف برای بیان کارکرد)

۲- کاربرد با تکرار بالا

۳- استفاده مکرر در موقعیت‌های دانشگاهی.

۴- هماهنگی با طرح مطالعه از لحاظ رابطه اجتماعی و وضعیت اجتماعی بین مصاحبان

۵- عدم وجه تشابه با کارکردها و ریز کارکردهای دیگر

شاید بعضی از این معیارها به‌نظر اختیاری یا ذهنی برسند. با این وجود، ایجاد چنین معیارهایی روش مناسبی در محدود کردن حیطه طرح آزمون می‌باشند. بعد از انتخاب کارکردها و ریز کارکردها باید عوامل اجتماعی موجود در ساختن پرسش‌های آزمون مورد توجه قرار گیرند.

۲-۴-۱۶ عوامل اجتماعی

۱-۲-۴-۱۶ موقعیت اجتماعی

موقعیت اجتماعی که عمل ارتباط در آن صورت می‌گیرد در تنوع زبانی مصاحبان از اهمیت والایی برخوردار می‌باشد. اما تنوع موقعیت‌های اجتماعی، مشکلاتی را در طرح آزمون بوجود می‌آورد. برای مثال، با استفاده از ساختارهای مختلف زبانی در دو موقعیت مختلف اجتماعی، کارکرد خاصی بعمل می‌آید. بنابراین برای طرح آزمون، شرایط اجتماعی خاصی همچون محیط‌های دانشگاهی روزمره و غیره باید تعیین شود.

۲-۲-۴-۱۶ رابطه اجتماعی

دو نقش کلی اجتماعی در عمل ارتباط وجود دارد (یعنی ارتباط دوست با دوست و یا ارتباط بیگانه با بیگانه)، هر سوال کارکردی برای این که شامل این دو رابطه اجتماعی شود باید بافت‌دار گردد.

بخصوص رابطه بین دو مصاحب در ارتباط است یا دوستانه است یا غیردوستانه. این رابطه با بیان این که آیا فرد با یک دوست در ارتباط است یا با کسی که اصلاً او را نمی‌شناسد، باید در محتوای هر سوال گنجانده شود. البته مسئله تعیین میزان دوستی یعنی میزان صمیمیت بین دو دوست نیز باید در طرح آزمون گنجانده شود. چنین سوال پیچیده‌ای که به کنش بین میزان دوستی و تنوع زبانی در اجرای یک کارکرد اشاره دارد، می‌تواند موضوع جالبی برای تحقیقات اجتماعی - زبانی باشد.

۳-۲-۴-۱۶ پایگاه اجتماعی

علاوه بر روابط اجتماعی مصاحبان، عامل اجتماعی دیگر یعنی وضعیت مصاحبان نیز باید مورد توجه قرار گیرد. می‌توان هر رابطه‌ای را برحسب وضعیت برابری یا نابرابری، طبقه‌بندی کرد. برای مثال، هر سوال می‌تواند به رابطه بین یک دانش‌آموز با دانش‌آموز دیگر (وضعیت برابر) یا یک دانش‌آموز با استاد (وضعیت نابرابر) مربوط باشد.

۳-۳-۴-۱۶ ملاک‌های عملکرد

تعیین مناسب‌ترین ساختارهای زبانی برای انجام یک کارکرد، کار آسانی نیست. زیرا در روش تدریس کارکردی، عملی کردن معیار برای یک جمله مناسب کار مشکلی می‌باشد. معیارهای تجربی متعددی وجود دارد که بطور ضمنی بر موفقیت فراگیران در میزان کارایی ارتباط اشاره دارند. ابهام چنین معیاری از لحاظ برقراری ارتباط موفق، نمره‌دهی آزمون را کامل مشکل می‌کند، زیرا چنین معیاری مستلزم قضاوت ذهنی می‌باشد. اگر چه معیار اصلی برقراری ارتباط مطلوب می‌باشد، میزان تناسب اجتماعی و یا صحت زبانی یک گفته که هر دو به یک میزان در فرآیند ارتباط تاثیر دارند، معین نشده است.

برای رفع چنین نقایصی می‌توان پرسش‌هایی با سه گزینه معنادار طرح کرد که هر یک هدف از پیش تعیین شده‌ای داشته باشند. این گزینه‌ها به چهار صورت می‌توانند وجود داشته باشند:

- ۱- مناسب از نظر اجتماعی و صحیح از نظر زبانی
- ۲- مناسب از نظر اجتماعی اما نادرست از نظر زبانی
- ۳- صحیح از نظر زبانی اما نامناسب از نظر اجتماعی
- ۴- نه مناسب از نظر اجتماعی و نه صحیح از نظر زبانی. این شکل پاسخ‌ها امتحان‌کنندگان را قادر می‌سازد تا به تفاوت توانایی زبانی امتحان‌شوندگان و آگاهی اجتماعی آنها پی ببرند.

۴-۴-۱۶ طرح تنه سؤال

اولین گام در طرح آزمون کارکردی انتخاب کارکردها و ریز کارکردها می‌باشد. گام بعدی ایجاد موقعیتهای واقعی می‌باشد که در آن این کارکردها اجرا می‌شوند. برای کسب اطمینان از واقعیت شرایط و مناسب بودن کارکردها در انعکاس روابط اجتماعی و وضعیت اجتماعی بین مصاحبان باید دقت لازم

به عمل آید. پس در جلسه‌ای متشکل از استادان و دانشجویان انواع مختلف تنه سؤالات مورد بررسی قرار می‌گیرند. هنگامی که اکثریت در مورد واقعی بودن موقعیتهای ذکر شده در تنه سؤالات موافقت کردند فرم نهایی سؤالات بایستی تهیه شود.

بهتر است سؤالات به شکل تشریحی باشند تا پاسخ‌های گویشوران بومی و غیربومی زبان را به دست آورد. در مرحله نهایی، نسخه اولیه آزمون با دستورالعملهای مناسب بایستی وارد مراحل پیش آزمون شود. برای روشن شدن مطلب، پرسشی به عنوان مثال ذکر شده است.

پرسش

فرض کنید داوطلب ورود به دانشگاهی می‌باشید. برای ورود به دانشگاه سه توصیه نامه از استاد مورد نیاز است. از استاد خود که دوستان نیز می‌باشد می‌خواهید یکی از نامه‌ها را بنویسد. به استاد خود چه می‌گویید؟

ویژگیهای پرسش

- ۱- موقعیت: دانشگاهی
- ۲- طرفین ارتباط: یک دانشجو و یک استاد
- ۳- نقش: انجام کاری
- ۴- ریزکارکردها: درخواست از دیگران برای انجام کاری
- ۵- رابطه اجتماعی: دوست با دوست
- ۶- وضعیت اجتماعی: نابرابر (دانشجو - استاد)

۵-۱۶ پیش آزمون

بدیهی است که به طور کلی طرح آزمون و خصوصاً طرح آزمون جدید کار آسانی نمی‌باشد. پیش آزمودن دقیق در تعیین اعتبار هر سؤال و همچنین اعتبار کل آزمون از اهمیت بسزایی برخوردار است. از آنجائیکه هر گزینه برای هدف خاصی طراحی شده است، مراحل مختلف پیش آزمون برای به دست آوردن جوابهای مختلف گزینه ضروری می‌باشد. اساساً سه مرحله پیش آزمون باید قبل از تکمیل آزمون تا مرحله اجرای عملی آزمون انجام بگیرد. در زیر هر مرحله به طور خلاصه مورد بررسی قرار گرفته است.

۵-۱-۶۱ مرحله ۱: پیش آزمودن گویشوران بومی زبان

این مرحله که به نظر می‌رسد مهمترین مرحله پیش آزمودن باشد بایستی در مورد گویشور بومی زبان انجام گیرد. افراد مورد مطالعه باید شامل دانشجویان فارغ‌التحصیل و دانشجویان در حال تحصیل مؤنث و مذکر از رشته‌های آموزشی مختلفی باشند.

آنها باید پاسخ مناسب به هر سؤال را بدهند (بدون هیچ محدودیتی در اندازه جوابها). هدف از انجام این مرحله کسب پاسخهای مناسب اجتماعی و صحیح زبانی برای هر پرسش می‌باشد.

۲-۵-۱۶ مرحله ۲: پیش آزمودن گویشوران غیربومی زبان

روش مشابه مرحله اول باید در مورد گویشوران غیربومی زبان با زمینه‌های مختلف زبان مادری‌شان انجام می‌شود. هدف اصلی این مرحله کسب و تعیین پاسخهای نامناسب ممکن برای هر سؤال می‌باشد. سپس این پاسخهای نامناسب با عملکرد گویشوران بومی مقایسه می‌شود تا دلیل انحراف آنها از هنجارهای گویشوران بومی مشخص شود. دلیل عدم تناسب دارای اهمیت می‌باشد، زیرا شاید برخی از آنها از نظر اجتماعی نامناسب باشند و برخی دیگر از نظر زبانی نادرست باشند و شاید بقیه پاسخها هم از نظر زبانی نادرست و هم از نظر اجتماعی نامناسب باشند.

۳-۵-۱۶ انتخاب گزینه‌ها

بعد از طبقه‌بندی پاسخ تمام سؤالات مهمترین مرحله انتخاب گزینه مناسب برای سؤال می‌باشد. برای اجتناب از هرگونه نظر شخصی و ذهنی و به‌جهت نظم و یکنواختی، پاسخهایی که به کرات توسط گویشوران بومی داده شده است، مناسبترین گزینه هر سؤال می‌باشند. بنابراین، در اولین گزینش پاسخهای مناسب اجتماعی و صحیح زبانی انتخاب می‌شوند به این پاسخ، پاسخ عملی گفته می‌شود.

مرحله بعدی تعیین پاسخهای نامناسب اجتماعی ولی صحیح زبانی از عملکرد گویشوران غیربومی می‌باشد. برای انجام چنین کاری تمام پاسخهای صحیح زبانی بدون صرفنظر از تناسب اجتماعی آنها باید در نظر گرفته شوند. سپس این پاسخها با پاسخهای داده شده توسط گویشوران بومی مقایسه شوند. این مقایسه به دو دلیل انجام می‌گیرد. اولاً، برای یقین از صحت زبانی پاسخها و دوماً، برای تعیین پاسخهای نادرست از نظر اجتماعی معیار انتخاب پاسخ نامناسب اجتماعی این است که آن پاسخ را هیچ گویشور بومی ارائه نداده باشد. بنابراین در دومین گزینش پاسخهای صحیح زبانی ولی نامناسب اجتماعی را می‌توان انتخاب کرد. این نوع پاسخ، پاسخ زبانی نامیده می‌شود.

سومین مرحله مشابه مرحله دوم می‌باشد، با این تفاوت که این بار تمام پاسخهای نادرست زبانی باید انتخاب شوند و با پاسخ گویشوران بومی مقایسه شوند. از میان این پاسخها، پاسخهایی که به هنجارهای گویشوران بومی از نظر مقبولیت اجتماعی نزدیک هستند انتخاب می‌شوند. این نوع پاسخ که از نظر اجتماعی مناسب می‌باشد اما از نظر زبانی نادرست، پاسخ اجتماعی نامیده می‌شود.

آخرین مرحله انتخاب، پاسخی است که هم از نظر اجتماعی و هم از نظر زبانی نسبت به هنجارهای گویشوران بومی زبان متفاوت می‌باشد. این پاسخها از میان عملکرد فراگیران بومی زبان انتخاب می‌شوند. هیچ یک از این پاسخها در عملکرد گویشوران بومی وجود ندارد. این گزینه به‌عنوان گزینه انحرافی می‌باشد.

۴-۵-۱۶ مرحله ۳: پیش آزمون فراگیران بومی و غیربومی زبان

اگر آزمونی به صورت چند گزینه‌ای طرح شده باشد، باید در مورد گویشوران بومی و غیربومی پیش آزموده شود. مهمترین هدفهای این مرحله عبارتند از:

الف) کسب اطمینان از این که گویشوران بومی پاسخهای کاربردی را انتخاب خواهند کرد.

ب) گویشوران غیربومی دستورالعملها و هدف آزمون را درک خواهند کرد.

بعد از آخرین بازنگریها و مرور بر روی سؤالات و گزینه‌ها شکل نهایی آزمون وارد مرحله اعتبار بخشی می‌شود.

۶-۱۶ سیستم نمره دهی

چنانکه قبلاً گفته شد، طبق شیوه تدریس کاربردی، پاسخ مناسب از نظر اجتماعی و صحیح زبانی، مطلوب‌ترین پاسخ بوده و نمره بیشتری نسبت به گزینه‌های دیگر دارند. بنابراین برای این نوع پاسخ در این آزمون دو نمره اختصاص داده می‌شود. به علاوه از آنجائیکه طراحان شیوه کاربردی در مورد رتبه‌بندی پاسخهای نسبتاً صحیح، یعنی این که آیا پاسخهای مناسب اجتماعی در ایجاد ارتباط مهمتر از پاسخهای صحیح زبانی می‌باشند یا نه جواب واضحی ارائه نداده‌اند، پس پیشنهاد می‌شود که به پاسخهای مناسب اجتماعی و صحیح زبانی ارزش یکسانی داده شود (هر کدام یک نمره). بالاخره برای چهارمین پاسخ هیچ نمره‌ای تعلق نمی‌گیرد. مزیت سیستم نمره‌دهی توزینی این است که امتحان‌شوندگان برای دانسته‌های خود (یا زبانی و یا اجتماعی) نمره می‌گیرند.

۷-۱۶ نتیجه گیری:

طبق ویژگیهای آماری گزارش شده برای آزمونهای کاربردی، این نوع آزمونها همچون آزمونهای دیگر دارای اعتبار و پایایی می‌باشند. اما آزمون کاربردی نسبت به سایر آزمونها دارای مزئیتهایی می‌باشد. اول این که، طرح آزمون براساس اصول معین نظریه‌ای می‌باشد که هدف از آزمون را روشن می‌سازد به این معنا که اعتبار محتوای چنین آزمونی خود به خود تضمین می‌شود زیرا طراحان آزمون قبل از طرح آزمون دقیقاً می‌دانند که چه چیز را مورد ارزیابی قرار دهند. دوم این که، برخلاف سایر آزمونها، پاسخ به پرسشهای آزمون براساس حدس یک یا چند طراح نمی‌باشد، یعنی شاید چیزی که به نظر طراح آزمون مناسب است، از نظر اکثریت گویشوران بومی صحیح نباشد. سوم این که چنین آزمونی می‌تواند دارای چندین هدف از جمله سطح‌بندی، مهارت و ارزیابی تشخیصی باشد.

علاوه بر این، آزمون کاربردی فاقد نقایص عمده آزمونهای مجزا است که عبارت بود از رده‌بندی مجزای رفتار تلفیقی زبان و همچنین این آزمونها معایب آماری و نظری آزمونهای تلفیقی را که شامل ابهام در مورد هدف و ارتباط سؤالات بود ندارد. آزمون کاربردی مهمترین ویژگی آزمونهای تلفیقی یعنی بافت مندی و مزیت مهم آزمونهای مجزا شکل یعنی عدم وابستگی متقابل سؤالات را در

یک روش آزمون سازی واحد گنجانده است. پس در آزمون کارکردی، سؤالات مستقل از یکدیگر بوده و به طور بالقوه بافت دار شده اند که به نظر می رسد حالتی ایده آل برای آزمونهای مهارت زبانی باشد. از این نظر آزمونهای کارکردی جایگزین مناسب و مکمل ضروری برای آزمونهای موجود می باشد. استفاده از آزمونهای کارکردی برای تعیین کاربرد مناسب زبان دارای نتایج مهمی برای طراحان آزمون و آزمون شوندگان می باشد. جنبه های ارتباطی زبان که با آزمونهای موجود قابل ارزیابی نمی باشند با آزمونهای کارکردی میسر می شود. این نکته مهمی است که متخصصان در زمینه آزمون باید توجه زیادی به آن بکنند.

فعالیت ها

- ۱- مزایا و معایب آزمون کارکردی را نسبت به سایر آزمون ها بیان کنید.
- ۲- آیا طرح آزمون کارکردی خواندن و امکان پذیر است.
- ۳- به نظر شما آزمون کارکردی در طرح آزمون های انگلیسی برای اهداف ویژه مؤثر است؟
- ۴- یک سوال کارکردی طرح کنید و ویژگیهای آنرا مشخص کنید.