المحاضرة الحادية عشر:

مؤشر الصعوبة، مؤشر التمييز، مؤشر ثبات المفردة

❖ **اهداف المحاضرة:**

يعد قياس مؤشر الصعوبة و التمييز و ثبات المفردة من اهم مكونات تحليل المفردات للمقاييس و بالأخص المقاييس المعرفية، و التي تعتمد على طرح الأسئلة سواء كانت ذات بدائل ثنائية، أو متعددة البدائل ...الخ و على هذا الأساس تم وضع معادلات لقياس هذه المؤشرات و كذا موازين تقويم لحذف او الاحتفاظ بالمفردات او البنود و على هذا الأساس كانت اهداف الدرس كما يلي :

- التعرف على مفهوم وطرق قياس مؤشر الصعوبة.
- التعرف على مفهوم وطرق قياس مؤشر التمييز.
- التعرف على طرق قياس ثبات المفردة.
- التعرف على معنى تصحيح أثر التخمين وطرق الوصول اليه.

1- مؤشر صعوبة المفردات:Item Difficulty Index

مؤشر الصعوبة او معامل الصعوبة يعتبر أحد المؤشرات المستخدمة ضمن عمليات تحليل البنود عندما تكون الاستجابة من نوع (صح / خطا)، ويقصد به تحديداً النسبة من أفراد العينة الذين تمكنوا من الإجابة على المفردة إجابة صحيحة. وفائدة مؤشر صعوبة المفردات تكمن في أنه يبين إلى أي مدى تناسب صعوبة المفردات مستوى الأفراد الذين عمل الإختبار من أجلهم.

1-1 طرق حساب مؤشر صعوبة المفردة :

أ- **الطريقة الأولى:** و هي ابسط الطرق، و هي مفضلة عندما تكون المجموعة صغيرة (عينة صغيرة)و يكون فيها معامل الصعوبة كما يلي :

معامل الصعوبة = (عدد الافراد الذين أجابوا إجابة صحيحة على السؤال / عدد الافراد الذين إجابو على السؤال)

$$P = c \div s \quad \text{.....(16) المعادلة}$$

مثلا لو كان لدينا 10 افراد أجاب منهم 7 على السؤال إجابة صحيحة و 3 إجابو إجابة خاطئة فان معامل الصعوبة هو (10/7) = 0.7 أي 70 % .

ملاحظة :يشير الخولي و عنان (1999) انه في حالة لم يجب احد الافراد عن السؤال فانه يحذف من المجموع الكلي.

في المثال السابق لو افترضنا انه احد الفراد ترك الإجابة و أجاب 7 منهم إجابة صحيحة و 2 إجابة خاطئة فانه تكون المعادلة كما يلي : (9/7) = 0.78 أي 78 % .

ب- **الطريقة الثانية** :في هذه الطريقة و التي يعتبرها البعض افضل من الطريقة الأولى و تتم وفقا لما يلي:

الخطوة 01 :ترتيب الإجابات في بطاقة التسجيل النهائية من اعلى الى الأدنى.

الخطوة 02 :تصنيف الإجابات في بطاقة التسجيل الجماعية الى ثلاث مجموعات فرعية كما يلي :

**المجموعة الأعلى : و تمثل 27 % من عدد الاجابات .

**المجموعة الأدنى : و تمثل 27 % من عدد الاجابات.

**المجموعة المتوسطة :و تمثل 46 % من عدد الاجابات.

وتضم المجموعة الأعلى المفحوصين الذين حققوا أعلى درجات على الاختبار، وتضم المجموعة الأدنى المفحوصين الذين حققوا أقل درجات على الاختبار، وتستخدم المجموعة الأعلى والمجموعة الأدنى كمجموعات متضادة تعكس الحد الأقصى للفروق في الأداء على الاختبار.

- يتوقف حجم كل مجموعة من المجموعتين (الأعلى والأدنى) على العدد الملائم بالنسبة لعملية التحليل الإحصائي، كما يتوقف على العدد الكلي للإجابات،و لكن هناك تقسيمات شهيرة لتحديد حجم كل مجموعة طرفية أهمها :

- طريقة كيلي Kelly 1939: حيث تتكون من 27 % كقيم طرفية لكل مجموعة .
- طريقة كيريتون Cureton 1957: حيث تتكون من 33 % كقيم طرفية لكل مجموعة .
- طريقة ايبل Ibel 1979: حيث تتكون من 25 % كقيم طرفية لكل مجموعة .
- طريقة ايبل Ibel 1979: حيث تتكون من 9 % كقيم طرفية لكل مجموعة .

الخطوة 03: حساب عدد الذين أجابوا إجابات صحيحة على كل فقرة من فقرات الاختبار بالنسبة للمجموعة الأعلىupper group .

الخطوة 4: حساب عدد الذين أجابوا إجابات صحيحة على كل فقرة من فقرات الاختبار بالنسبة للمجموعة الأدنى Lower group .

$$P=\frac{C_1+C_2}{n1+n2}$$ .....(17) المعادلة

حيث :

P =مؤشر الصعوبة

C_1 = عدد الإجابات الصحيحة في المجموعة الاعلى.

C_2 = عدد الإجابات الصحيحة في المجموعة الادنى.

n_1 = عدد الأفراد في المجموعة الأعلى. n_2 = عدد الأفراد في المجموعة الأعلى.

مثال :تم تطبيق اختبار معرفي على عينة تكونت من186 لاعب ،و بعد نهاية الاختبار تم ترتيبهم وفقا لإجاباتهم على الاختبار من الأعلى الى الأدنى كالتالي :

الفقرات / المجموعات	ع1	ع2	ع3	ع4	المجموع
المجموعة الأعلى 27 %	28	2	6	14	50
المجموعة الأدنى 27 %	24	8	1	17	50

مؤشر صعوبة الفقرة (ع1)

$$P_1 = \frac{28+24}{50+50} = 0.52$$

و بنفس الطريقة يتم حساب مؤشر صعوبة باقي الفقرات:

*مؤشر صعوبة الفقرة (ع1) = 0.52 أي 52 % *مؤشر صعوبة الفقرة (ع2) = 0.10 أي 10 %

*مؤشر صعوبة الفقرة (ع3) = 0.07 أي 7 % *مؤشر صعوبة الفقرة (ع4) =0.31 أي 31 %

2-1 **دلالات مؤشر الصعوبة** :تتراوح صعوبة المفردات بين صفر ، و الواحد الصحيح (0 – 1)، حيث كلما ارتفعت قيمة المؤشر و اقتربت الى 1 دل ذلك على سهولة الفقرة ،و كلما انخفض و اتجهت نحو الصفر دل ذلك على صعوبة الفقرة ،و الجدول التالي يبين :

جدول رقم (10)دلالات مؤشر صعوبة المفردات

مؤشر الصعوبة	%	الدلالة	التقويم
0.8 فاكثر	80% فأكثر	سهلة جدا	تستبعد الفقرة (غير قادرة على التمييز)
0.79 – 0.71	من 71 % الى 79%	سهلة	يمكن قبول الفقرة اذا كان لها معامل تمييز مقبول
0.70–0.30	من 30 % الى 70 %	متوسطة	الفقرة جيدة
0.29–0.20	من 20 % الى 29 %	صعبة	يمكن قبول الفقرة اذا كان لها معامل تمييز مقبول
اقل من 0.2	أقل من 20%	صعبة جدا	تستبعد الفقرة (غير قادرة على التمييز)

المصدر: (رضوان، 2006)

و الأمر الحاسم في تحديد مستوى صعوبة المفردات هو الغرض الذي من أجله سيتم تطبيق الإختبار فلو كان الغرض من أجل الاختيار من بين المتقدمين في برنامج الدراسات العليا أو لو كان للإختيار من بين متقدمين إلى وظيفة ففي هذه الحالة لابد من أن يكون مستوى صعوبة المفردات عالياً حتى نضمن إمكانية التمييز بين المتقدمين . ومن الإشكاليات المتعلقة بصعوبة الفقرات هو عمليات التخمين التي يقوم بها المفحوصون خاصة في إختبارات الصح والخطأ وإختبارات الإختيار من متعدد ، ذلك أن الأمر يتطلب تعديل مستوى الصعوبة ففي أسئلة الصح والخطأ قد يكون مستوى الصعوبة 0.50 عندما يخمن الطلاب في الإجابة وفي هذه الحالة يتطلب الأمر تعديل مستوى الصعوبة لنأخذ في الاعتبار عملية التخمين وسيكون مستوى الصعوبة الأفضل في هذا الوضع 0.75 أي في الوسط بين 0.50 والواحد الصحيح .

2 مؤشر القدرة التمييزية للبنود : Item Discrimination Index

القدرة التمييزية للبنود تعتبر إحدى الخصائص التي يسعى مطور الإختبار للتأكد منها عندما يقوم بعملية تحليل البنود كخطوة من خطوات بناء الاختبار والمقصود بالقدرة التمييزية الدرجة التي تتمتع بها الفقرة من حيث التمييز الدقيق والصحيح بين المفحوصين في السمة أو الخاصية التي يقيسها الإختبار–بمعنى قدرة الفقرة على ان تميز بين الافراد الذين يحصلون على درجات منخفضة في الصفةأو السمة التي تقيسها كل الفقرات المقياس أو الاختبار–

ومن أجل حساب مؤشر القدرة التمييزية الفقرات لابد من العناية والإهتمام بنوع القياس المستخدم هل هو من النوع المتصل Continuous أم من النوع المنفصل Dichotomous بالإضافة إلى إفتراض التوزع الإعتدالي للصفة أو الخاصية المقاسة ، كما أن من الأمور ذات الأهمية العلاقة بين الصعوبة والقدرة التمييزية للفقرات فالبنود الصعبة جداً لا تميز بين المفحوصين كما أن البنود السهلة جداً لا تميز بينهم كذلك فالبنود الصعبة يفشل المفحوصون في الإجابة عليها ، بينما البنود السهلة يتمكن كل المفحوصين أو غالبيتهم من الإجابة عليها .

2-1 طرق حساب معامل التمييز :

أ- الطريقة الأولى :طريقة المجموعتين المتطرفتين *The Extreme Group Method*

و فيها يتم تقسيم المفحوصين الى مجموعتين متساويتين في العدد وفقا للدرجة الكلية للاختبار ، المجموعة العليا (U)و تضم أصحاب الدرجات الجيدة (المتميزين) ،و المجموعة الدنيا (L)و التي تضم

أصحاب الدرجات الدنيا (الضعيفة) ، و يمكن استخدام احدى الطرق التي ذكرناها سابقا(مثل دريقة كيلي Kelly 27% في كل مجموعة طرفية).

ثم يتم تطبيق المعادلة التالية:

$$d = \frac{U-L}{N}$$ المعادلة (18).....

حيث

d= مؤشر التمييز للفقرة

U=عدد الإجابات الصحيحة على الفقرة في المجموعة العليا (المتميزة)

L = عدد الإجابات الصحيحة على الفقرة في المجموعة الدنيا (الضعيفة)

N= عدد المفحوصين في احدى المجموعتين (العليا او الدنيا)

- **دلالات مؤشر التمييز وفق الطريقة الاولى:**

- يتراوح مؤشر التمييز بين +1 و −1 .
- فعندما يحصل عدد كبير من المفحوصين في المجموعة الأعلى على إجابات صحيحة على الفقرة نقول ""ان **الدرجة التمييزية للفقرة إيجابية** " بمعنى ان للفقرة مؤشر تمييز إيجابي Positive discrimination، أي ان اتجاه التمييز للفقرة هو نفس اتجاه الدرجة الكلية للاختبار.
- يشير التمييز السلبي للفقرة Positive discrimination الى ان أكثر المفحوصين في المجموعة الدنيا يحصلون على إجابة صحيحة للفقرة.
- فاذا كان لفقرة الاختبار القدرة على التمييز بصورة جيدة، فإن ذلك يعني أن المفحوصين الذين يحققون درجات عالية (مرتفعة) على الاختبار سوف يجيبون على هذه الفقرة إجابة صحيحة، عكس المفحوصين الذين لهم درجات منخفضة على الاختبار، وتعد الفقرة غير القادرة على التمييز فقرة ضعيفة البناء.poorly Construction
- تحذف الفقرات التي مؤشر تمييزها سالب (أي عدد المفحوصين في المجموعة الدنيا الضعيفة أجابو على الفقرة بطريقة أفضل من المجموعة العليا الجيدة).
- قد حدد ايبل 1979 (Ibel) مؤشرات التمييز للفقرات في الاختبارات معيارية المرجع كالتالي :

جدول رقم (11) مؤشرات تمييز الفقرات في الاختبارات معيارية المرجع

قيمة مؤشر التمييز	التقويم	الحكم
0.40 فما أكثر	لها قدرة عالية جدا على التمييز	فقرة جيدة جدا
0.30 الى 0.39	لها قدرة جيدة على التمييز	فقرة جيدة
0.20 الى 0.29	لها قدرة متوسطة على التمييز	قد تحتاج الفقرة الى تعديل
0.10 الى 0.19	لها قدرة ضعيفة على التمييز	تحتاج الى مراجعة كبيرة
اقل من 0.10	ليس لها قدرة على التمييز	تستبعد الفقرة

المصدر (dela Peña Jr، 2011؛ رضوان، 2006)

و اذا أخذنا كل من مؤشر الصعوبة و معامل التمييز و دمجنا تقديراتهم يمكننا الحصول على الجدول التالي:

جدول رقم(12) تصنيف العبارات وفق معامل الصعوبة و التمييز في الاختبارات معيارية المرجع

معامل التمييز / مؤشر الصعوبة	اقل من 0.10	0.10 الى 0.19	0.20 الى 0.29	0.30 الى 0.39	0.40 فما أكثر
أقل من 0.20	تستبعد	تستبعد	تستبعد	تستبعد	تستبعد
من 0.2 الى 0.29	تستبعد	تستبعد	مراجعة كبيرة	مراجعة	مراجعة
من 0.29 الى 0.70	تستبعد	مراجعة	مراجعة	تقبل	تقبل
من 0.71 الى 0.79	تستبعد	مراجعة كبيرة	مراجعة	تقبل	تقبل
0.80 فأكثر	تستبعد	تستبعد	مراجعة كبيرة	مراجعة	مراجعة

المصدر:من اعداد الباحث اعتماد عن عدة مراجع

اليك المثال التالي: تم تطبيق اختبار معرفي على عينة تكونت من186 لاعب ،و بعد نهاية الاختبار تم ترتيبهم وفقا لإجاباتهم على الاختبار من الأعلى الى الأدنى كالتالي :

الفقرات / المجموعات	ع1	ع2	ع3	ع4	عدد الافراد
المجموعة الأعلى 27 %	28	2	20	31	50
المجموعة الأدنى 27 %	12	8	9	17	50

المطلوب استخرج معامل تمييز الفقرات

D1=(28−12)/50=0.32 D2=(2−8)/50=−0.12

D3=(20−9)/50=0.22 D4=(31−10)/50=0.42

- من خلال النتائج يتبين
- ✓ الفقرة ع1 لها قدرة لها قدرة جيدة على التمييز.
- ✓ الفقرة ع2 تحذف.
- ✓ الفقرة ع3 لها قدرة متوسطة على التمييز.
- ✓ الفقرة ع4 لها قدرة متوسطة على التمييز.

❖ ملاحظة01 : ان مستوى القدرة التمييزية المطلوب ليس محدداً في قيمة معينة بل إن من العوامل ما يجب أخذه في الإعتبار عند إختيار البنود حسب عاملي صعوبتها وتمييزها في الغالب الاختبارات معيارية المرجع يفضل أن يكون معامل تمييزها عالياً فالاختبارات التحصيلية لابد من أن تميز بين الطلاب أما الإختبارات التي يراد منها إستثارة دافعية الطلاب فالواجب أن تتضمن فقرات ذات تمييز عالي وأخرى ذات تمييز منخفض ، أيضاً الإختبارات التي من شأنها أن تكشف عن الإستعدادات وكذا إختبارات الإختيار الوظيفي كلها لابد من أن يكون معامل تمييز فقراتها عالياً حتى يمكن الإختيار والتصنيف للأفراد .

❖ ملاحظة 02 :ان الغرض من تطبيق الاختبار هو الذي يفرض إما فقرات عالية في معامل تمييزها أو منخفضة ويرى (Mehrens & Lehman) أن معامل تمييز بحدود 0.40 سيكون مرض بشكل عام لكنهما يؤكدان على أن الفقرات منخفضة معامل التمييز أو التي معاملات تمييزها سالبة الإشارة لابد لها من مراجعة وتعديل حتى تكون مناسبة لغرض الإختبار.

ب- **الطريقة الثانية "الطرق الارتباطية"**: اطلق ليندكويست Lindquist 1951 على هذه الطريقة اسم "مؤشر تمييز الفقرة عن طريق الاتساق الداخلي " ،حيث تعتمد هذه الطريقة على تحليل العلاقة الارتباطية بين الدرجة على الفقرة و الدرجة الكلية للاختبار

""the correlation between performance on the item and performance on the total test""

فالعلاقة بين الفقرة و الدرجة الكلية للاختبار تظهر كيف تقيس الفقرة بشكل جيد الوظائف التي يقسها الاختبار نفسه، حيث يبين مؤشر الصدق المدى الذي تمتد اليه فقرة محددة في التمييز بين المفحوصين الذين يظهرون فروقا حادة في الوظائف التي يقسها الاختبار ككل،حيث يرى بعض خبراء القياس أنه يمكن النظر الى ارتباط الدرجة الكلية للفقرة على انه مؤشر جيد للتمييز ،حتى انه يمكن

اختيار الفقرات على أساس الارتباطات الكلية للفقرة ،اذا كانت نسبة الارتباطات الموجبة بين 0.15 و 0.75 .

ويستخدم لقياس الارتباط هناك معاملات شائعة أشهرها معامل الارتباط ثنائي التسلسل النقطي و بعض المعاملات الأخرى :

- معامل الارتباط ثنائي التسلسل النقطي "معامل الارتباط الثنائي الأصيل" (The point biserial correlation)
- معامل الارتباط ثنائي التسلسل (Biserial correlation)
- معامل فاي Phi.
- معال الارتباط الرباعي tetrachoric

➢ **حساب مؤشر الصعوبة باستخدام معامل الارتباط ثنائي التسلسل النقطي (The point biserial correlation)**

و يسمى أيضا بمعامل الارتباط الثنائي الأصيل و يستخدم لحساب العلاقة الارتباطية بين متغيرين احدهما اسمي ثنائي أصلا (dichotomous) ،و المتغير الثاني مستمر (فتري) يتوزع توزيعا طبيعيا ،و يعتبر حالة خاصة من معامل ارتباط بيرسن.

- تتراوح قيمته بين −1 و +1 ، و كلما ارتفع دل ذلك على ان الفقرة لها معامل تمييز جيد
- إذا كانت هذه القيمة سلبية أو منخفضة، فيجب حذف العنصر من الاختبار. كلما اقتربت قيمة المؤشر من 1.0، كان العنصر أفضل
- نقوم أيضًا بتصنيف وتلخيص أسئلة الاختبار التي يحتمل أن تكون جيدة أو سيئة من خلال الارتباط ثنائي التسلسل النقطي في تقرير مصفوفة العناصر على النحو التالي:

"ضعيف" <0.15 "مقبول"[0.15 −0.3] "جيد"> 0.30.

لاحظ أن العناصر الأسهل، مثل تلك التي تمت الإجابة عليها بشكل صحيح بنسبة 90٪ أو أكثر، لا تبدو عادةً عناصر جيدة في مؤشر التمييز. إذا حصل 90٪ من المتقدمين للاختبار على عنصر ما بشكل صحيح، فهناك تباين ضئيل جدًا في الأداء بحيث يكون هناك ارتباط كبير مع مجموع درجات الاختبار. وبالمثل، إذا كانت العناصر صعبة للغاية بحيث تمت الإجابة عليها بشكل صحيح بنسبة 10٪ أو أقل من المتقدمين للاختبار، فهناك مساحة صغيرة جدًا لإظهار الارتباط بين العناصر وإجمالي درجات الاختبار.

و لحساب معامل التمييز فانه يتم حساب معامل الارتباط بين الفقرة التي تكون الإجابة عليها (بصح /خطا) و والأداء الكلي للاختبار وفقا للمعادلة التالية :

$$r_{pbis}=\left[\frac{\overline{Y_c}-\overline{Y_t}}{S_t}\right]\sqrt{\frac{P_x}{1-P_x}}$$ ..... المعادلة (19)

r_{pbis} = معامل الارتباط النقطي أو مؤشر التمييز .

$\overline{Y_t}$ = متوسط الدرجة في الاختبار ككل لجميع المفحوصين.

$\overline{Y_c}$ = المتوسط الحسابي للاختبار ككل للمفحوصين الذين اجابوا اجابة صحيحة على الفقرة.

S_t = الانحراف المعياري لدرجات الاختبار ككل.

P_x = نسبة الافراد الذين أجابوا إجابة صحيحة على الفقرة من بين جميع المفحوصين.

ويمكن كتابة المعادلة السابقة كما يلي:

$$r_{pbis}=\left[\frac{\overline{Y_c}-\overline{Y_t}}{S_t}\right]\sqrt{\frac{P_x}{q_x}}$$ .....المعادلة (20)

حيث:

P_x =نسبة الافراد الذين أجابوا إجابة صحيحة على الفقرة .

q_x = نسبة الافراد الذين أجابوا إجابة خاطئة على الفقرة.

مثال:في احدى الاختبارات المعرفية أجاب 58 % من افراد العينة على الفقرة رقم 05، وبلغ المتوسط الحسابي للاختبار ككل بالنسبة للأفراد الذين أجابوا إجابة صحيحة على الفقرة رقم 05 هو 57.60، والمتوسط الحسابي للاختبار ككل لجميع افراد العينة هو 54.30، اما الانحراف المعياري للاختبار بالنسبة لجميع افراد العينة على الاختبار ككل هو 9.7.

المطلوب استخراج مؤشر الصعوبة باستخدام الارتباط المتسلسل النقطي؟

من خلال المعطيات يتبين ان: $\overline{Y}$ = 54.30 $\overline{Y_1}$ = 57.60 S_y = 9.7 P_x = % 58

بالتعويض في المعادلة الخاصة بحساب r_{pbis}:

r_{pbis} =[(57.60 − 54.30)/9.7]

$$r_{pbis}=\frac{57.60-54.30}{9.70}\sqrt{\frac{0.58}{1-0.58}}=\frac{23.3}{9.70}\sqrt{\frac{0.58}{0.42}}=(0.34)(1.17)=0.40$$

وبالتالي فان مؤشر التمييز للفقرة رقم 05 هي 0.40 وتعبر عن معامل تمييز جيد.

3- مؤشر ثبات الفقرات: item reliability index

يمكن حساب مؤشر ثبات الفقرات في الاختبارات التي لها فقرات ثنائية الإجابة (صح/ خطا) كما يلي : ثبات الفقرة= الانحراف المعياري للفقرة X معامل الارتباط بين الفقرة و الاختبار.

المعادلة التالية :

المعادلة (21).....$ire=(S_i)\ (r_{pbis})$

r_{pbis} = معامل الارتباط النقطي بين الفقرة و الاختبار (معامل التمييز)

S_i = الانحراف المعياري للفقرة و الذي يمكن حسابه كما يلي : $S_{i=}\sqrt{pq}$.

p = هي معامل الصعوبة (نسبة كامل الافراد الذين أجابوا إجابة صحيحة على الفقرة)

q = تمثل نسبة الافراد الذين أجابوا إجابة خاطئة على الفقرة اي (q=1−p)

و بالتالي يمكن كتابة المعادلة السابقة كما يلي :

المعادلة (22).....$ire=\sqrt{pq}(r_{pbis})$

و يتم استخدام مؤشر ثبات البنود هنا (item reliability index) كوسيلة للاختيار بين بندين (فقرتين) في بناء النموذج (الاختبار) فإذا كان هناك فقرتين لهما قيم تمييز متشابهة، ولكن احدى الفقرات لها انحراف معياري (تباين) أعلى من الاخرى، فسيكون لهذا البند ثبات أعلى وسيساهم بشكل أكبر في الثبات التقييم. مع تساوي كل شيء آخر، قد يقرر مطور الاختبار الاحتفاظ بالبند ذو الثبات الأعلى ويحتفظ بالبند الأقل ثبات في بنك الفقرات كاحتياطي.

❖ **ملاحظة** :في حالة استخدام معامل الارتباط ثنائي التسلسل النقطي فانه يمكننا اعتباره معامل صدق اتساق داخلي للفقرات .وهو ما يراه (سعد عبد الرحمان) قائلا "و يعتبر معامل تمييز البند أو قدرته دليلا عل صدقه و خاصة اذا كان الامر ينطوي على مقارنة طرفي القدرة التي يقيسها ،حيث ان صدق الاختبار انما يعتمد على صدق و حداته او بنوده و قدرتها على التمييز ،و من ثم فان حساب درجة تمييز كل بند سوف يهيئ الطريقة للحصول على اختبار صادق في حالة ارتفاع معاملات التمييز...و بحساب درجة صدق البند او قدرة البند على التمييز فان ذلك يعني اننا نحقق الاساسيات العامة للاختبار ،و خاصة فيما يتعلق بقدرة الاختبار على التمييز بين طرفي القدرة التي يقيسها"

و يؤكد علام (2000) انه "ربما نتخذ معامل التمييز دليلا على صدق المفردة مع الاخذ في الاعتبار التحقق من صدق محتوى المفردة قبل تضمينها في الاختبار.(علام، 200، صفحة 377)

- الا اننا نؤكد على أن كل هذا اجتهادات و الافضل استخدام محك خارجي لقياس مؤشر صدق الفقرة.

4- تصحيح أثر التخمين:CORRECTING FOR GUESSING

في بعض الأحيان تتأثر نتائج الاختبار بما يسمى بالصدفة او اثر التخمين، و عليه لا بد قبل أخذ نتائج الاختبار يجب ان نصححها من أثر التخمين او الصدفة في الإجابة او الاختيار.

4-1 **أسباب اللجوء الى تصحيح اثر التخمين** :هناك من يرى ضرورة تصحيح قيمة صعوبة مفردات الاختبار من أثر التخمين العشوائي؛ وذلك للحصول على تقدير نسبة عدد الأفراد الذين يجيبون إجابة صحيحة عن كل مفردة استنادا إلى معلوماتهم الفعلية واستبعاد من يجيبون إجابة صحيحة عن طريق التخمين العشوائي .ويستند هذا التقدير إلى افتراض أن اختيار الفرد إجابة غير صحيحة لمفردة اختبارية يرجع إلى افتقاره للمعلومات التي تتطلبها الإجابة الصحيحة ، وليس إلى خطأ معلوماته أو سوء فهمه، وكذلك افتراض أن الفرد الذي لا يعرف الإجابة الصحيحة يرى أن جميع البدائل التي تشتمل عليها المفردة يمكن أن تمثل الإجابة الصحيحة. فإذا قبلنا هذين الافتراضين، فإننا نتوقع توصل نسبة معينة من الأفراد الذين لا يعرفون الإجابة الصحيحة إلى هذه الإجابة بمحض الصدفة، وهذه النسبة هي متوسط نسب عدد الأفراد الذين يختارون أحد البدائل الأخرى أي الإجابات الخطأ ويمكن الحصول على تقدير نسبة عدد الذين يجيبون إجابة صحيحة عن المفردة استنادا إلى معلوماتهم الفعلية بطرح هذه النسبة من النسبة الكلية لعدد الأفراد الذين يجيبون عن المفردة إجابة صحيحة.

و عموما يمكننا ان نلجأ الى تصحيح أثر التخمين عندما:

- في حالة الاختبارات المعرفية التي تتطلب السرعة في الأداء، والتي تعرف باسم اختبارات السرعة، لكونها اختبارات تعتبر السرعة في الأداء عاملا مهما، لذلك يتوقع ان يحدث التخمين عند الاستجابة على مثل هذه الاختيارات.
- في الاختبارات المعرفية التي تتميز بدرجة عالية من الصعوبة وتكون محدودة بزمن.
- قد لا تكون هناك حاجة لتصحيح أثر التخمين عندما ينتهي جميع المفحوصين من الاختبار قبل انتهاء الزمن المخصص له، لان هذا يعني أن أثر التخمين في الإجابة كان ضعيفا للغاية لعدم وجود أية ضغوط مرتبطة بالزمن قد تؤدي الى التخمين.
- يلزم تصحيح أثر التخمين في الاختبارات التي تصمم بحيث لا يستطيع المفحوص الإجابة على جميع وحداتها خلال الزمن المقرر للاختبار.

- في حالة الفقرات ذات البدائل الثنائية (صح / خطا) تتأثر بعوامل التخمين أكثر منه بالنسبة لاي نمط بخر من الاختيارات، حيث كلما زاد عدد الاختيارات قل تأثير التخمين .

4-2 أسباب عدم اللجوء الى تصحيح أثر التخمين:

فالبعض يرى أنه ليس من الضروري إجراء هذا التصحيح، وحجتهم في ذلك ما بلى:

- لا يتغير ترتيب الأفراد فيما بينهم إذا أجرينا تصحيحا لدرجاتهم الكلية في الاختبار من أثر التخمين.
- يؤدي تصحيح درجات الأفراد الذين لا يلجأون إلى التخمين إلى خفض درجاتهم الكلية في الاختبار، بينما لا يحدث ذلك لمن اعتاد على التخمين في إجابته.
- يعتمد تصحيح الدرجات من أثر التخمين على فرضية أن الإجابة الصحيحة إما تكون معلومة أو غير معلومة، في حين أن بعض الأفراد يلجأون إلى التخمين استنادا إلى معلومات جزئية لديهم.
- تصحيح الدرجات من أثر التخمين ربما يؤدي إلى الافتراض الخاطئ بأن الدرجات الناتجة هي الدرجات الحقيقية للأفراد في الاختبار.

لذا يقترح اصحاب هذا التوجه بعض الإرشادات التي يمكن أن تسهم في يمكن تقليل أثر التخمين بقدر كبير كالتالي

4-3 بعض الارشادات لتقليل أثر التخمين: يمكن تقليل أثر التخمين بقدر كبير كالتالي

- تحذير المختبرين من التخمين العشوائي في إجاباتهم عن مفردات الاختبار.
- السماح بوقت كاف لكل فرد لكي يتمكن من محاولة إجابة كل مفردة من مفردات الاختبار.
- تضمين البدائل التي تشتمل عليها مفردات الاختيار من متعدد معلومات صحيحة جزئيا أو أخطاء شائعة تتعلق بالمجال الذي تقيمه المفردات .

4-5 طرق تصحيح أثر التخمين :و يمكن تصحيح اثر التخمين على قيمة الدرجة باستخدام المعادلة التالية :

$$SAC = R - \frac{W}{k-1} \quad \text{.....المعادلة(23)}$$

حيث :

SAC = الدرجة بعد التصحيح / R= عدد الإجابات الصحيحة

W= عدد الإجابات الخاطئة. / k = عدد البدائل لكل فقرة .

أما معامل الصعوبة بعد التصحيح فيمكن حسابه كما يلي :

$$P_a = P_b - \left(\frac{1-P_b}{k-1}\right)$$ المعادلة (24).....

حيث :

P_a = معامل الصعوبة بعد التصحيح / P_b = معامل الصعوبة قبل التصحيح

k = عدد الاختيارات (عدد البدائل)

مثال: في احدى البحوث كانت هناك اربع فقرات حيث تم تقسيمهم وفق لاجاباتهم الى مجموعتين عليا تتكون من 27 فرد ،و دنيا تتكون من 27 فرد حيث :

الفقرة الأولى (ف1) تتكون من بديلين (صح/خطا) حيث تحصلنا في المجموعة العليا على 17 إجابة صحيحة و في المجموعة الدنيا 8 خاطئة.

القرة الثانية (ف2) تتكون من أربع بدائل و تحصلنا حيث تحصلنا في المجموعة العليا على 17 إجابة صحيحة و في المجموعة الدنيا 8 خاطئة.

- المطلوب احسب قيمة معامل الصعوبة قبل تصحيح أثر التخمين و بعد تصحيح اثر التخمين.

أ- معامل الصعوبة قبل تصحيح أثر التخمين:

نقوم تطبيق المعادلة الخاصة بمعامل الصعوبة و قد تحصلنا على النتائج التالية :

*معامل صعوبة الفقرة الأولى = 0.46 أي 46 %

*معامل صعوبة الفقرة الثانية =0.46 أي 46 %

نلاحظ أن معامل الصعوبة لم يختلف في السؤالين بالرغم من أن احدها تتكون من بديلين و الأخرى من 4 بدائل و سوف نرى ماذا سيحدث ان قمنا بتصحيح أثر التخمين:

ب- حساب معامل الصعوبة بعد تصحيح أثر التخمين و هذا بعد تطبيق المعادلة 09 :

- معامل صعوبة الفقرة الأولى بعد التصحيح = -0.08
- معامل صعوبة الفقرة الثانية بعد التصحيح = 0.28

نلاحظ من خلال النتائج بعد تصحيح اثر التخمين ان العبارة الأولى أصبح معامل صعوبتها سالبا و بالتالي تحذف ،اما العبارة الثانية فاصبح 0.28 .

و هذه النتيجة تؤكد ان العبارات ثنائية البدائل تتأثر بصوبرة كبيرة باثر التخمين في مقابل العبارات التي لديها بدائل متعددة ابر من اثنان .

5- الاستجابات المتعددة في الاختبارات المعرفية:

قد نلجا الى استخدام طريقة البدائل المتعددة في أسئلة الاختبار المعرفي مثلا اربع بدائل في عوض بديلين ،و يجب ان ننتبه عند بناء الاختبار المعرفي و تطبيقه على العينة الأولية (التجريبية/الاستطلاعية) الى هذه البدائل و يطلق على **الاستجابة الصحيحة اسم الاختيار الصحيح** ،و على الاستجابات الخاطئة اسم : **المظللات او المشتتات** ،و تستمد اسمها من وظيفتها لكي تظلل المفحوصين المشككين في الإجابة الصحيحة للفقرة، يمكن الإبقاء على البديل الذي اختاره ما بين (2 الى 3 %) من عدد المفحوصين الذين تم اختبارهم و الا يمكن ان يستبعد هذا البديل.

حيث هناك شروط نذكر منها:

1- البديل الذي لم يختاره أي فرد من المجموعة العليا والدنيا (0 اختيار) يستبعد من البدائل الخاصة بالفقرة.

2- البديل غير الصحيح و الذي تختاره المجموعة الدنيا أكثر من المجموعة العليا يسمى بدلي مظلل او مشتت.

3- البديل الصحيح و الذي تختاره المجموعة العليا اكثر من المجموعة الدنيا .

4- البديل الخاطئ الذي تختاره المجموعة العليا اكثر من المجموعة الدنيا لا يصلح ان يكون بديل مشتت او مظلل و يستبعد.

5- لا بد ان تحقق الفقرة مؤشر صعوبقة وتمييز مناسب لابقائها.

مثال :في احد البحوث تم وضع مجموعة من البدائل امام أحد الفقرات و تحصلنا على النتائج التالية حيث ان البديل (ب2) هو الإجابة الصحيحة

البدائل / المجموعات	ب1	ب2	ب3	ب4
المجموعة الأعلى (30)	5	22	0	5
المجموعة الأدنى(30)	4	07	0	17

المطلوب :

1- تقويم الاستجابات (البدائل) الخاصة بهذه الفقرة .

2- استخرج معامل صعوبة الفقرة قبل و بعد تصحيح أثر التخمين .

3- استخرج معامل تمييز الفقرة و هل يتم ابقاؤها .

1- تقويم الاستجابات :

- نلاحظ أن البديل (ب1) عبارة عن إجابة خاطئة وقد اختاره (5) من الفئة العليا وهو اكبر من الذين اختاروه من الفئة الدنيا و الدين عددهم (4) و بالتالبي لا يصلح كاستجابة مشتتة او مظللة.
- نلاحظ ان البديل (ب3) لم يختره أي فرد من المجموعتين العليا و الدنيا و بالتالي لا تستبعد لعدم فعاليتها.
- نلاحظ ان البديل (ب4) عدد الذين اختاروه من الفئة العليا (5) افراد وهو اقل من الذين اختاروه من الفئة الدنيا (17) فرد ،و بالتالي تصلح كاستجابة مظللة .

2- استخراج معامل الصعوبة قبل تصحيح أثر التخمين :هو 0.48 ،اما معامل الصعوبة بعد التصحيح = 0.31

3- معامل التمييز = 0.50

❖ **أسئلة التقويم:**

1- كيف يتم حساب مؤشر الصعوبة ومؤشر التمييز وفق طريقة المجموعتين المتطرفتين؟
2- متى يتم استخدام الارتباط الثنائي الأصيل في حساب مؤشر تمييز الفقرة ؟
3- متى يتم استبعاد الفقرة ؟
4- كيف يتم تصحيح أثر التخمين لمعامل الصعوبة ؟
5- متى يعتبر البديل مشتت او مظلل ؟
6- ماهي اهم طرق صياغة البنود في الاختبارات المعرفية ؟