ثبات أختبار المواقف لتقييم أداء الطالب المعلم فى ضوء نظرية إمكانية التعميم [1]

إعداد/ جوزيف سمير وهيب كامل[2]

مدرس مساعد بقسم علم النفس التربوى

كلية التربية – جامعة السويس

الملخص العربي:

تهدف الدراسة الحالية إلى التحقق من ثبات درجات أختبار المواقف لتقييم أداء الطالب المعلم فى ضوء نظرية إمكانية التعميم، ومعرفة مكونات التباين الاكثر تاثيراً على ثبات درجات تقييم الأداء بواسطة ذلك الاختبار، ومعرفة أفضل الطرق لتحسين إجراءات القياس المستقبلية وأكثرها فعالية فى تصميمات القياس للوصول لدرجات قابلة للتعميم، وقد طبقت الدراسة الحالية على (٢٧١) طالب وطالبة من طلاب الفرقة الرابعة بكلية التربية– جامعة السويس حيث طبق عليهم أختبار المواقف، وكشفت النتائج وصول أختبار المواقف لدرجة جيدة من معامل إمكانية التعميم النسبى والمطلق مما يدل على ثبات درجات أختبار المواقف، وإن أكبر مكونات التباين تاثيراً على معامل إمكانية التعميم النسبى والمطلق تمثلت فى تفاعل طالب– أختبار، وتفاعل طالب– أختبار– فترة. فى حين جاءت باقى مكونات التباين الآخرى ضعيفة ولا تؤثر على معاملات إمكانية التعميم، كما أوضحت النتائج أن زيادة عدد الابعاد وزيادة عدد مرات التطبيق تؤدى إلى زيادة معاملات إمكانية التعميم النسبية والمطلقة أفضل من زيادة عدد المُصححين.

مقدمة:

نتيجة للإنتقادات المتعددة التي أبرزتها البحوث والدراسات التربوية فى السنوات الأخيرة لإستخدامات التقييمات التقليدية للإختبار؛ فقد تزايد الإهتمام بالتقييمات البديلة القائمة على الأداء والتى جاءت كرد فعل للتقييمات التقليدية التى تعتمد علي إختيار الإجابة والتدريس من أجل الأختبار وقياس معارف ومهارات منعزلة، إضافة إلي تركيزها على قياس العمليات المعرفية الدُنيا، والتقييم المباشر المبني على المعارف دون الأداء. لذا أنتقلت بعدها الممارسات التربوية

---

[1] تم استلام البحث في ٢٥/ ١٠/ ٢٠٢٢ وتقرر صلاحيته للنشر في ١/ ١٢/ ٢٠٢٢

[2] ت: ٠١٢٢٨١٣٧٧٣٢ Email: joseph.samir@suezuni.edu.eg

نحو إستخدام تقييمات الأداء فى مجالات التكوين والعمل مما يسمح بتعزيز المعارف والمهارات والاتجاهات ذات القيمة فى التربية (صلاح الدين علام، ٢٠٠٠؛ ٢٠٠٤).

وحيث أن المعلم يمثل العامل الرئيسى فى أى نظام تعليمى والقوة الفاعلة فى المنظومة التعليمية، ويعد عنصراً فعالاً فى تحقيق أهداف التربية الأمر الذى جعل الاهتمام به مدخلاً من المداخل الأساسية لإصلاح التعليم، لذا فأن تقييم أداء المعلم عملية هامة وضرورية ويجب أن تستند على تحليل عمل مهنة المعلم حتى تكون عملية وموضوعية وتفيد المعلم نفسه، وتفيد الأطراف التربوية الآخرى ذات الصلة بتدريبه وتأهيله؛ لذا يجب أن تواكب ما يستحدث من تطبيقات تقنية حتى يستطيع المعلم أن يؤدى أداءاً تدريسياً فعالاً ومبدعاً تظهر نتائجه في مخرجات التعلم (Kerry, 2015).

ولقد استُخدمت النظرية الكلاسيكية للقياس لفترة طويلة من الزمن فى عملية بناء الاختبارات وتطبيقها وتحليل وتفسير نتائجها، الا أنها لم تتمكن من معالجة بعض القضايا فى القياس والتقويم التربوى، وقد وجهت لهذة النظرية الكثير من الانتقادات مثل معالجتها لمفهومى الصدق والثبات نظراً لانها لاتميز بين أخطاء القياس المتعددة، وإنما تعطى قيمة تقديرية كلية لمصادر أخطاء القياس المتعددة المتعلقة بإحدى صور الاختبار المستخدمة بواسطة فاحص معين، وفى ظروف معينة. لذلك تتعدد أساليب تقدير الثبات، ونتيجة لذلك أوضح (فؤاد ابو حطب، ١٩٩٢،٣٢١) إن علماء القياس النفسى والتربوى قاموا بتطوير وتعديل نظريات القياس حتى تتلائم مع مختلف أدوات القياس التى تُستخدم في الحكم والتقييم؛ للوصول إلى دقة تقدير الصدق والثبات لتلك الأدوات، وتحقيق موضوعية القياس.

وتعد نظرية إمكانية التعميم من النظريات المعاصرة المستخدمة لتقدير صدق وثبات تقييمات الاداء نتيجة لما فرضتة تقييمات الأداء من تعدد أبعاد القياس والواقعية وإمكانية التعميم ووجود العديد من المصادر المحتلمة المختلفة لاخطاء القياس لذا وجب الرجوع الي طرق إحصائية أكثر مرونة وتلائماً ودقة لمعالجة تلك الأدوات بدلاً من النظرية الكلاسيكية التي أصبحت غير كافية ومجدية. (Brennan, 2010).

مشكلة البحث:

بالرغم من حدود النظرية الكلاسيكية للاختبارات فى معالجتها لأخطاء القياس إلا إنها ساهمت بشكل مباشر فى ظهور نظرية إمكانية التعميم، وقدمت إطاراً فكرياً ساعدت على تطوير نظرية إمكانية التعميم بواسطة إستخدام طرق تحليل التباين فى تقدير ثبات القياسات السلوكية. لذلك لا

يمكن أعتبار نظرية إمكانية التعميم كبديل للنظرية الكلاسيكية ولكن يمكن اعتبارها كأمتداد لها لتلافى عيوبها، فهى عبارة عن مفهوم يجمع بين طرق تحليل التباين ونظرية القياس الكلاسيكية بأسم إمكانية التعميم. حيث أُعتبر برينان (Brennan,2001) أن النظرية الكلاسيكية وتحليل التباين كأبوين لنظرية إمكانية التعميم حيث قدموا حلول متعددة عن كيفية تجزئة مصادر التباين المتعددة فى القياس وتقدير مكونات تباين الدرجات وتفاعلاتها المتعددة في نفس الوقت وبالتالي التحرر من فكرة خطأ أحادى البعد غير مميز.

وتؤكد نظرية إمكانية التعميم على أن هناك مصادر متعددة للخطأ فى موقف القياس منها مثلاً خصائص المفحوصين، وخصائص الفاحصين، ووقت تطبيق الأختبار، ونوع فقرات الأختبار و... غيرها، وإن هذه الأخطاء قد تحدث معاً بصورة متزامنة أثناء القياس أو منفردة. فنظرية إمكانية التعميم ترى أن الدرجة الملاحظة يمكن تقسيمها الي درجة شاملة و واحد أو أكثر من مكونات الخطأ(Shavelson & Webb, 1991) لذا فلن يكون من المناسب أن يعتمد الباحث علي درجة واحدة أُخذت من موقف واحد ووفق ظروف دون ظروف آخرى، وأعتمدت على مصدر (بُعد) واحد للخطأ دون بقية الابعاد مثلما الحال في النظرية الكلاسيكية (Eason, 1989 فى: صبرى محمود، ٢٠١٦).

وعلى الرغم من محاولة نظرية الاستجابة للمفردة من معالجة أوجه القصور فى النظرية الكلاسيكية ومحاولة كلاً من النظريتين أن تعالج تباين أخطاء القياس معالجة مختلفة، إلا أن كليهما ينظران إلى مصدر هذة الاخطاء نظرة ضيقة ومحدودة (صلاح الدين علام، ٢٠٠٠، ٦٩٩).

ومن الناحية التطبيقية لا يوجد أى نموذج أو نظرية متكاملة حتى الآن غير نظرية إمكانية التعميم يمكنها أن تعالج فى آن واحد مختلف مصادر الخطأ فى موقف القياس، وتوفير معلومات دقيقة عن كل مصدر من مصادر التباين الاخرى وتفاعلاتها مع بعضها البعض، إضافة إلى توفير طرق مختلفة لتعميم وتحسين إجراءات القياس لم تكن متوفرة فى كلً من النظرية التقليدية والاستجابة للمفردة (طباع فاروق، ٢٠٢٠).

كما أشار كاردينت وآخرون (Cardinet et al.,2010) إن نظرية إمكانية التعميم لا تقدم وسائل لتقدير ثبات قياسات أنجزت من قبل فحسب، ولكن أيضاً تقدم معلومات حول مساهمة الاخطاء لإستخدامها فى تحسين إجراءات القياس فى المستقبل. لذلك تقوم دراسات القرار على إستخدام المعلومات المتحصل عليها من دراسات إمكانية التعميم لتصميم أفضل للقياس السلوكى.

كما أن التركيز في الماضى على المهارات الاساسية أدى إلى افتقار الطلبة إلى المهارات

العقلية العليا، وإن الاختبارات التى تقتصر على الورقة والقلم والاختيار من متعدد فقط كان لها آثار سلبية كثيرة، حيث أنها ركزت توجه المعلمين والطلبة نحو التدريب على هذة الاختبارات والاعداد لها، وبذلك عززت الممارسات التربوية الآلية وجعلت التعلم سطحى وسلبى مما أسهم فى إعاقة الإصلاح الفعلى للتعليم حيث تم التركيز على درجات الاختبارات بدلاً من تحسين تعلم الطلبة؛ مما أدى إلى إصلاحات تستند إلى تقويم بديل يركز على التقييم القائم على الاداء نظراً لتعدد وتنوع مستويات النواتج التعليمية (صلاح الدين علام، ٢٠٠٤).

فإستخدام مقاييس الأداء يسهم فى عمليات ضبط وتوثيق نتائج القياس يزيد من الدقة فى الاداء، وتحديد الفروق بشكل واضح، ويوفر الوقت والجهد والمال في عمليات التقويم، ويزيد من الموضوعية ويقلل من التحيز، لذا إكتسب أهمية فى شتى العلوم فجميع العلوم تسعي لتطوير أساليب موضوعية دقيقة لقياس الظواهر المتعلقة بها من أجل فهم هذة الظواهر وتفسيرها، والتنبؤ بالعلاقات القائمة بين متغيراتها، ومحاولة ضبطها والتحكم بها(Cary, 1994).

ونظراً لشيوع إستخدام التقييمات التقليدية مثل الاختيار من متعدد والصواب والخطأ و... غيرها فإن تقدير الخصائص السيكومترية لها نالت اهتمام الباحثين لكثير من الوقت، وبنفس الطريقة إنتقل الاهتمام أيضاً نحو تقييمات الأداء خلال العقديين الماضيين لقياس جودتها الفنية خاصة ما يتعلق بصدق وثبات تلك التقييمات نظراً لاختلافها عن التقييمات التقليدية وتعدد أبعادها وإعطائها حرية أكبر للطلاب فى الإجابة عن المهام التي تعكس إدماج معارف ومهارات متعددة، وواقعية سياقها، وإستدعائها لمهارات تفكير عليا؛ مما يتطلب إعادة النظر فى مفاهيم الصدق والثبات الكلاسيكية (يسرى زكى، ٢٠١٦).

حيث أن استخدام النظرية الكلاسيكية لتقدير صدق وثبات تقييمات الأداء أصبحت غير كافية لوجود العديد من المصادر المحتلمة المختلفة لأخطاء القياس لذا وجب الرجوع إلي طرق إحصائية أكثر مرونة وتلائماً ودقة لمعالجة ذلك بأستخدام مفاهيم ومبادئ نظرية أمكانية التعميم(Brennan, 2001; 2010 ; Suen & lei, 2007)، حيث تبحث نظرية إمكانية التعميم في العلاقات بين أخطاء القياس المختلفة مع بعضها البعض بشكل متزامن وليست بشكل منفصل كما تفعل النظرية الكلاسيكية للقياس؛ فهى لاتستطيع أن تبحث أخطاء القياس الناتجة من المصححين والفترات وصور المهام و ... غيرها فى وقت واحد معاً كما تفعل نظرية إمكانية التعميم (Shavelson& webb, 1991; Atilla&Ezel, 2010).

لذا عند فحص الدراسات في الأدب النفسى والسيكومترى، يُلاحظ أن نظرية التعميم تُستخدم

غالبًا فى تقدير ثبات تقييمات الأداء(Lee& Kantor, 2007; Yılmaz&Gelbal, 2011; Yılmaz&Başbaşa, 2015)، وقد قام هوانج (Huang, 2009) بجمع ودراسة نتائج أكثر من ٥٠ دراسة أشتملت علي ١٣٠ مجموعة بيانات مستقلة خلال الاعوام من ١٩٨٠ إلى ٢٠٠٦ تناولت تقييمات الاداء من خلال تطبيقات نظرية إمكانية التعميم فى مجال تقييمات الاداء، وقام بالتوصل إلى نتائج كمية فى عدة أبعاد من حيث طريقة التقييم (إختبار ورقة وقلم، وأنواع آخرى)، ومجالات الموضوع (رياضيات، والعلوم، والاصغاء، والقراءة، والكتابة، واللغة الاجنبية، والاداء العسكرى، والاداء الطبى).

وبالرغم من أعتماد أغلب تقييمات الاداء بمختلف أنواعها وتخصصتها علي نظرية إمكانية التعميم، فقد أعتمدت بعض الدراسات العربية والاجنبية الآخرى مثل دراسة كلاً من (Demet & Erkut, 2018)، (سامية مداح، ٢٠١٧)، (ياسر عبدالحافظ وعبدالله سالم، ٢٠١٣)، (Maria & Maria, 2013)، (Mohamed, 2010) بالرغم من حداثتهم فى تقييم أداء المعلم – فى مختلف تخصصاتهم– على إستخدام نظرية القياس التقليدية فى القياس بالرغم من عدم فاعليتها فى مقاييس تقييم الأداء دون اللجوء الي نظرية إمكانية التعميم رغم قدرتها وأفضليتها فى معالجة تقييم الأداء، كما أن أغلب تلك الدراسات تقوم على إستبيانات تقليدية لمجموعة من المهام والكفايات للمعلم دون الاعتماد على تحليل عمل مهنة المعلم.

وأستخدم كارى (Carrie, 2013) نظرية إمكانية التعميم لتحديد مصادر التباين فى أداة الملاحظة التجريبية المصممة لتقييم فعالية معلم التربية الخاصة، وقد خلصت النتائج إلى ضرورة وجود محكمين مختلفين مثل (مديري المدارس وأعضاء هيئة التدريس بالجامعة ... الخ)، وضرورة إستخدام أحجام أكبر لعينات الاوجه(محكمين/مواقف) لزيادة دقة القياس للأداة المستخدمة.

كما أوضح كلاً من (Bell et al., 2012; Hill et al., 2012) أن النموذج الكامل للحكم علي جودة المعلم يتضمن مجموعة من الأبعاد/المكونات وهى: كل مستويات المفردات، المقيمين، مناسبات القياس عبر الايام (الوقت)، أيام التقييم، الاقسام، كما أوضح (Bakker et al., 2008) أن زيادة عدد المقيمين وزيادة مهام التقييم وتنوعها تؤدى إلى معاملات أكثر تعميم عند تقييم أداء المعلم.

كما أوضحت العديد من الدراسات التى تناولت تقييم الأداء بأستخدام نظرية إمكانية التعميم مثل دراسة كلاً من (Shavelson et al., 1993; Lane et al., 1996; Webb et al., 2000;

Gao& Brennan, 2001; Nie et al., 2007;Guler& Gelbal, 2010; Taylor& Pastor, 2013) وداسة كلاً من (صبرى محمود، ٢٠١٦)، ودراسة (طباع فاروق، ٢٠٢٠) أن أكبر مصدر للتباين يعود إلى تفاعل الطالب مع الاختبار مقارنة بباقي مكونات التباين الاخرى التي جاءت ضعيفة، وهذا يتفق مع التحليل الذى أجراة هوانج (Huang, 2009) حول الدراسات التى أهتمت بتقييمات الاداء فى تخصصات مختلفة و أوضحت تلك الدراسات أن نسبة كبيرة من مصدر الخطأ يرجع إلى تفاعل الطالب مع المهمة أو الاختبار.

فى حين اظهرت دراسات آخرى (Chen et al., 2007; Lee& Kantor, 2007; Gebril, 2009) بأن أكبر مصدر للتباين يرجع إلى تفاعل الطالب مع الاختبار مع المُصحح الممزوج بالاخطاء العشوائية غير المقدرة، بالإضافة إلى دراسة كازانوفا وديماس (Casanova& Demeuse, 2011) التى كشفت أن اكبر مصدر للتباين راجع إلى المصححين، وتفاعل المصححين مع الاختبارات.

وفيما يتعلق بمستوى ثبات الاختبارات المستخدمة فى تقييم أداء الطلاب، فقد تباينت نتائج الدراسات حيث أظهرت بعض الدراسات (Lane et al., 1996; Lee& Kantor, 2007; Guler& Gelbal, 2010) بلوغ درجات الاختبارات مستويات مقبولة من الثبات، إلا أن البعض الآخر من الدراسات مثل (Shavelson et al., 1993; Nie et al., 2007; Taylor& Pastor, 2013) كشفت عن ضعف فى ثبات الاختبارات.

ومن ناحية آخرى أثبتت معظم الدراسات (Nie et al., 2007; Lee& Kantor, 2007; Guler& Gelbal, 2010; Taylor& Pastor, 2013)، وداسة كلا من (صبرى محمود، ٢٠١٦)، و(طباع فاروق، ٢٠٢٠) أن زيادة عدد الاختبارات تساهم أكثر فى زيادة معاملات إمكانية التعميم مقارنة بزيادة عدد المصححين، وهو على عكس ما توصلت إلية دراسة كازانوفا وديماس (Casanova& Demeuse, 2011) التى كشفت أن زيادة عدد المقدرين أفضل من زيادة عدد المهمات فى زيادة معاملات إمكانية التعميم، كما أوضح هوانج (Huang, 2009) أنة يمكن إستخدام طرق آخرى لخفض مكونات التباين تمثلت فى إدماج أكبر عدد من الأبعاد فى موقف القياس، وإستخدام تصميمات متقاطعة بدلاً من تصميمات متداخلة، وإدماج الفترة كبعد من أبعاد القياس.

إن الاختلاف فى نتائج الدراسات السابقة حول أختلاف مصادر تباين الخطأ المؤثرة على ثبات درجات تقييم الطلاب بالإضافة إلى مدى تحقيق الاختبارات لمستويات مقبولة من الثبات، وقلة

الدراسات فى البيئة العربية والمصرية التى تتناولت نظرية إمكانية التعميم؛ يكون مبرراً لإجراء البحث الحالى؛ لذلك سوف تهتم الدراسة الحالية بالاجابة على مجموعة من التساؤلات المرتبطة والمنقسمة إلى تساؤلات ترتبط ارتباطاً وثيقاً بتصميمات نظرية امكانية التعميم التى تهتم بتفسير وتقدير مكونات التباين المؤثرة على دقة القياسات، وتساؤلات ترتبط ارتباطاً وثيقاً مع دراسات القرار التى تهتم بتحسين إجراءات القياس للوصول إلى مستويات مقبولة من إمكانية التعميم فى المستقبل.

وبالتالى يمكن صياغة مشكلة البحث فى الاسئلة التالية:

١- ما حجم التباين الذى تساهم بة كل مصدر من مصادر التباين (الأبعاد) فى التباين الكلى والتى تؤثر على تقدير ثبات تقييم أداء الطالب المعلم على أختبار المواقف ؟

٢- ما تأثير زيادة كل مصدر من مصادر التباين (الأبعاد) المؤثرة على ثبات درجات أداء الطالب/المعلم على أختبار المواقف؟

أهداف البحث:

يهدف البحث الحالي إلى :

١- معرفة مكونات التباين الاكثر تاثيراً على ثبات درجات أختبار المواقف لتقييم أداء الطالب المعلم.

٢- معرفة أفضل طرق تحسين إجراءات القياس وأكثرها فعالية فى تصميمات القياس، عن طريق زيادة عدد الابعاد لمصادر التباين للحصول على درجات قابلة للتعميم .

أهمية البحث:

يكتسب البحث الحالي أهميتة من العديد من المنطلقات منها :

١- المساهمة فى توضيح وفهم مفاهيم ومبادئ نظرية امكانية التعميم، وتقييمات الاداء، ودلالة وأهمية كل منهما من الناحية النظرية والتطبيقية فى تقدير دقة القياسات وإجراءات تحسين عمليات القياس للكثير من الباحثين لعدم الالمام بها حتى الآن.

٢- المساهمه فى تقييم الاداء المهنى للمعلم بإستخدام أداة مبنية علي تحليل عمل مهنة المعلم ونظرية إمكانية التعميم تكون بمثابة أختبار قبول للطلاب بكليات التربية لضمان الحد الادنى من الأداء المطلوب للقبول بكليات التربية، ويمكن أن تستخدم كرخصة لمزاولة مهنة التدريس.

مصطلحات البحث:

نظرية إمكانية التعميم Generalizability Theory(GT)

نظرية إحصائية لتقييم صدق وثبات القياسات السلوكية، وتركز النظرية على مصادر أخطاء القياس المتنوعة حيث تقوم بفصلها وتقديرها وترتيبها فى نفس الوقت (Salkined, 2008, 436)، تعرفها (Urbina, 2004, 138) بأنها تقدير بديل للثبات، وامتداد للنظرية الكلاسيكية للقياس تستخدم طرق تحليل التباين فى تقدير التأثيرات الكلية للمصادر المتعددة لتباين الخطأ فى درجات الاختبار فى آن واحد.

تقييم الاداء Performance Assessment

هو أن يظهر المتعلم بوضوح، أو يبرهن، أو يقدم أمثلة، أو تجارب أو نتاجات أو... غير ذلك، تتخذ دليلاً على تحقيقة مستوى تربوياً أو هدفاً تعليمياً معيناً. وتتطلب مهام الاداء إجراء العمليات والتوصل إلى نتاجات، وأهم ما يميز هذة المهام أنها مباشرة ووظيفية وواقعية أى تمثل مواقف حياتية فعلية خارج نطاق الصف المدرسى كما أنها لا تتطلب بالضرورة الورقة والقلم (صلاح الدين علام، ٢٠٠٤، ١٠٥).

الإطار النظرى والدراسات المرتبطة:

أولاً: نظرية إمكانية التعميم

يري كرونباخ Cronbach وراجارتنام Rajaratnam وجليزر Glaser أن درجات الاختبار أو أى درجات ملاحظة تُعد عينة من نطاق شامل للدرجات الممكنة وتتباين وتختلف ظروف جمع تلك الملاحظات وفقاً لأبعاد متعددة، ومثال على ذلك تطبيق اختبار معين مرة واحدة، أو تطبيقه وإعادة تطبيقه واشتماله على عينة ممثلة من المفردات دون بقية العينات المستمدة من النطاق الشامل، وتطبيقه تحت ظروف معينة دون باقى الظروف، وهذا التباين فى الأبعاد المتعلقة بأداة القياس وما تتضمنة من مفردات وظروف تطبيقها وخصائص مجموعة الأفراد التى يطبق عليها الاختبار تؤدى إلى مشكلات فى عمليات القياس وإمكانية تعميم نتائجها، وقد أقترح كرونباخ وزملائه عام ١٩٧٢م نظاماً لتقدير تأثير كل من هذه الأبعاد فى تباين الدرجات الملاحظة وتقدير دقة الدرجة الحقيقية من الدرجات الملاحظة التى تتأثر بجميع أبعاد عملية القياس وإمكانية تعميم تلك الدرجات الملاحظة فى ضوء هذه الأبعاد المتعددة عند القياس، وقد عرف ذلك النظام باسم نظرية إمكانية التعميم Generalizability Theory أو نظرية الثبات متعددة الأوجه

Multifaceted Reliability Theory أو نظرية عينة المجال Domain Sampling Theory فهى نظرية تبرز أهميتها واسهاماتها فى توسيع نطاق استخدامات كل من مفهومي الصدق والثبات وتكاملهما (صلاح الدين علام، ٢٠٠٠، ١٧١).

أسباب ظهور نظرية إمكانية التعميم:

١- يري كرونباخ ومعاونيه أن ضعف النظرية الكلاسيكية تكمن فى التبسيط الشديد لمفهومي الصدق والثبات مما جعل الباحثين يستخدمون هذين المفهومين استخداماً إلىاً، مما أدى إلى غموض معناهما الحقيقى واستخدامهما الفعلى؛ لذا فقد اهتم كرونباخ ومعاونيه بمراجعة وتعميق مفهومي الصدق والثبات ومعالجتهما معالجة أكثر شمولاً من منظور متعدد الأبعاد (Nicholas, 2016).

٢- لا تسمح الطرق المختلفة لتقدير الثبات بالنظرية الكلاسيكية مثل إعادة التطبيق، والصور المتكافئة، والاتساق الداخلى لصانع القرار من التحكم فى أخطاء القياس الأخرى التى قد تؤثر على الدرجة الملاحظة أثناء استخدام إحدى هذه الطرق مما يؤدى إلى غموض فى تفسير صدق وثبات درجات المقاييس فى معناهما واستخداماتهما الفعلىة؛ فتعدد الطرق المستخدمة فى تقدير الثبات يعبر عن عدم قدرة النظرية الكلاسيكيةعلى التحكم فى الأخطاء المتعددة التى يمكن أن تؤثر فى عملية القياس، كما لا تتميز طرق قياس الثبات بالمرونة الكافية لتتناسب مع مشكلات الثبات التى تظهر فى جميع الاختبارات العقلية (Crocker& Algina, 2006) كما تبين أيضا أن تطوير النظرية الكلاسيكية لطرق متعددة فى تقدير الثبات تُحدد الثبات بشكل مختلف لا يتماثل مع التعريف النظرى له (Webb et al., 2006).

٣- على الرغم من أن النظرية الكلاسيكية ونظرية الاستجابة للمفردة تعالجان تباين أخطاء القياس معالجة مختلفة؛ إلا أن كليهما ينظر إلى مصدر هذه الأخطاء نظرة ضيقة ومحدودة ولا يمكنهما التمييز بين أخطاء القياس؛ فالمعلومات أو البيانات التى تستمد سواء من العلوم الطبيعية أو العلوم السلوكية تكون مشوبة بدرجة معينة من الخطأ، وقد تكون بعض أنواع أخطاء القياس غير متضمنة فى عملية القياس وذلك يضع قيداً على نطاق ظروف القياس التى يمكن التعميم عليها؛ فنظرية إمكانية التعميم تعتمد على تحديد ظروف القياس التى تسهم فى الخطأ لكى يمكن تعميم القياس عليها (صلاح الدين محمود، ٢٠٠٠، ٦٩٩).

أهمية نظرية إمكانية التعميم

نظرية إمكانية التعميم تُعرف بنظرية متعددة الأبعاد فهى تزودنا بمعلومات مفصلة عن مختلف

مصادر الخطأ فى درجات الاختبار عن طريق جمع ملاحظات متعددة لنفس مجموعة الأفراد فى كل المتغيرات المستقلة التى تساهم فى تباين الخطأ فى القياس مثل المهام، والمصححين، والفترات الزمنية، وصيغ مهام الاختبار، وأيضاً التفاعلات بين الابعاد وبعضها البعض (طباع فاروق، ٢٠٢٠).

كما أن نظرية إمكانية التعميم لم تهتم فقط بكيفية بناء أدوات قياس تكون نتائجها متكافئة فى ظروف ومواقف اختبارية مختلفة، وإنما اهتمت أيضاً باستخدامات أدوات القياس فى صنع قرارات معينة، فنظرية إمكانية التعميم تميز تحليلات الثبات من منظورها إلى مرحلتين مختلفتين، المرحلة الأولى تسمى بدراسة إمكانية التعميم تهدف إلى تحديد المصادر أو الأبعاد المتعددة لأخطاء القياس وتقييم مقدار التباين فى الدرجات المحصلة من كل مصدر أو بعد من أبعاد القياس، وتسمي المرحلة الثانية بدراسة القرار وتهدف إلى تحديد الثقة فى درجات القياس وإمكانية تصميم إجراءات قياس أكثر فعالية (Meyer, 2010).

وتفترض نظرية إمكانية التعميم أن الدرجة الشاملة هي بديل للدرجة الحقيقية فى نظرية القياس الكلاسيكية، وتقسم خطأ القياس إلى مصادر خطأ متعددة حيث يعتبر التمييز بين مصادر الأخطاء عند القياس من المساهمات الأساسية لنظرية إمكانية التعميم (Bertrand& Blais, 2004).

يرى شافىلسون، وويب (Shavelson& Webb ,1991) أن قوة نظرية إمكانية التعميم تكمن فى إمكانية تقدير مختلف مصادر القياس بشكل مستقل فى تحليل واحد، إضافة إلى تمكين صانع القرار من تحديد المصادر الكامنة التى تسهم فى تباين الدرجات للحصول على درجة عالية من ثبات القياس.

المفاهيم الأساسية لنظرية إمكانية التعميم

إن ما تتطلبة تلك النظرية من تصورات فكرية وأساليب إحصائية مختلفة عن باقى نظريات القياس الآخرى تتطلب بعض المفاهيم الاساسية لفهمها، ومن المفاهيم الاساسية للنظرية:

١) النطاق الشامل: Universe Domain

هو نطاق يتضمن شروط جمع الملاحظات أو الحصول على القياسات حتى نتمكن بعد ذلك من إمكانية تعميم هذه القياسات من موقف إلى آخر ومن عينة الأفراد إلى عينة آخرى؛ فهو الهيكل أو الإطار الذي يحدد خصائص كل بعد متضمن فى عملية القياس (Meyer, 2010).

٢) الدرجة الشاملة: Universe Score

هى القيمة المتوقعة لدرجات الفرد الملاحظة والتى يحصل عليها فى مختلف المواقف التى

تنتمى إلى النطاق الشامل المحدد من قبل الباحث. وللفرد أكثر من درجة شاملة على حسب الأبعاد التى يحتوى عليها النطاق الشامل. كما أن الدرجة الشاملة هي درجة تقابل الدرجة الحقيقية فى النظرية الكلاسيكية (صلاح الدين علام، ٢٠٠٠، ٧٠٢).

٣) البعد/ الوجه/ مصادر الخطأ: Facet

هو عبارة عن المصادر التى قد تؤثر على موقف القياس مثل الاختبارات والمصححين والاهداف والفترات وصيغ الاختبار و... غيرها التى يمكن أن تختلف من ملاحظة إلى آخرى وبذلك يمكن أن تؤثر على درجة الفرد الملاحظة (Brennan, 2010).

٤) المستويات/ الشروط: Levels/ Conditions

كل بعد من الأبعاد يُشكل من مجموعة من المستويات التى تعبر عن مختلف شروط أو ظروف القياس فى هذا البعد، مثل عدد الأفراد الذين يطبق علىهم الاختبار، وعدد المقدرين الذين يقدرون أداء الطلاب، ومختلف فترات أداء الطالب للمهام و... هكذا (Cardinet et al., 2010, 11)

٥) تحليل التباين: Analysis of Variance

هو أسلوب يستخدم لعزل تباين متغير واحد من تباين المتغيرات الآخرى بواسطة تجزئة التباين الكلى لمجموعة الملاحظات التى ترجع إلى عوامل معينة مثل الجنس، ومجموعات المعالجة الآخرى (Everitt& Skrondal, 2010).

٦) دراسات إمكانية التعميم: Generalizability Study (G–Study)

هي دراسات تهتم بتحديد درجة تكافؤ نتائج الاختبارات التى نحصل عليها فى ظروف اختبارية مختلفة وتحت شروط متباينة؛ أى أنها دراسات تستخدم فى جمع معلومات وبيانات تفيد فى تقدير مكونات تباين القياسات التى نحصل علىها بطريقة ما من الطرق، وهذا يفيد فى تطوير هذه القياسات من خلال محاولة تقليل أثر تلك التباينات على أخطاء القياس (صلاح الدين علام، ٢٠٠٠، ٧٠٤).

٧) معامل إمكانية التعميم: Coefficient Of Generalizability

هو النسبة بين تباين الدرجة الشاملة إلى تباين الدرجة الملاحظة، وبالتالى يمكن اعتبار معامل إمكانية التعميم بديلاً عن معامل الثبات الذي نحصل علىة باستخدام النظرية الكلاسيكية للقياس، وكل اختبار أو مقياس يكون له أكثر من معامل إمكانية تعميم على حسب الأوجه التى يتم أخذها بعين الاعتبار فى دراسة إمكانية التعميم (صلاح الدين علام، ٢٠٠٠، ٧٠٤).

٨) دراسات القرار: Decision Study (D-Study)

هى دراسات تهدف الوصول إلى أفضل إجراء قياس يتمتع بالثبات فى موقف معين اعتماداً على المعلومات المتوفرة من دراسات إمكانية التعميم، وبالتإلى فإنها تجيب على التساؤل كيف يمكن الوصول بثبات الاختبار إلى أفضل مستوى، وكيفية تحسين إجراءات القياس بالاعتماد على تقارير دراسة إمكانية التعميم وتقديرات خطأ التعميم الخاص بالتصميم المستخدم فى الدراسة (طباع فاروق، ٢٠١٩).

الخصائص السيكومترية للاختبارات فى إطار نظرية إمكانية التعميم

يري كلا من كرونباخ Cronbach وجليسر Gleser أن الصدق والثبات مفهومان مترابطان، ويمكن أن يندرجا تحت أسم مقياس إمكانية التعميم، ويكمن الفرق الرئيسي بينهما فى الأبعاد التى نود التعميم عليها (صلاح الدين علام، ٢٠٠٤) لذا تعد نظرية إمكانية التعميم من نظريات القياس المعاصرة التى ساهمت فى إبراز التكامل بين مفهومي الصدق والثبات، ويمكن استغلال نتائج ثبات القياسات السلوكية فى التأكد من صدق أنظمة التقييم خاصة إذا تعلق الامر بصدق التكوين الفرضي، ويتم ذلك من خلال تحليل مكونات التباين التى تفسر تباين الأداء فى الاختبارات مما يساعد على تحديد النطاق الشامل للشروط المقبولة.

كما أشار بيركر وآل (Berker& Al, 1993)، و موس (Moss, 1992) أن التمييز بين الصدق والثبات فى نظرية إمكانية التعميم أصبح ضئيلاً، حيث يمكن الاخذ بالاعتبار فى تقييم الأداء العلاقات بين المهام كأدلة لكلاً من الصدق والثبات، كما أوضح شيفلسون و ويب (2009) Shavelson& Webb أنة يمكن اعتبار بعض أبعاد موقف القياس أبعاداً للصدق فى ضوء نظرية إمكانية التعميم، وركزا على مجالات المحتوي المتعددة وأنواع المهام (اختيار من متعدد، مفتوحة، يدوية،...) كأبعاد للصدق تكون ضمنية فى تصميم دراسة إمكانية التعميم.

وقد أوضح أيضاً شافيلسون وآخرون (1993) .Shavelson et al التكامل بين الصدق والثبات فى إطار نظرية إمكانية التعميم فى تقييم الأداء من خلال فحص تغير معاينة مختلف طرق تقديم المهام دليل على الصدق التقاربي للاختبار. لذا يمكن أعتبار مختلف إجراءات القياس التى تعني بقياس نفس السمة وتعطي نفس الصورة دليلاً على الصدق التقاربي، ويمكن أعتبار تعميم عينة مهام أو بنود الاختبار على مجال واسع دليلاً على صدق المحتوى، ويمكن اعتبار مقدار التغير داخل القياس إذا كانت الفترة تمتد إلى أسابيع أو أشهر أو أبعد من ذلك إلى سنه دليلاً على الصدق التنبؤي (Shavelson& Webb, 2009) فالصدق والثبات مفهومان مترابطان يختلفان

فقط فى الابعاد التى يود صانع القرار التعميم عليها.

ثانياً: تقييم الأداء

يطلق الباحثين على هذا النوع من التقويم عدة أسماء، وهى جميعها تمثل مرادفات لمعنى واحد، ومن أمثلة ذلك التقويم الشامل Comprehensive Assessment، والتقويم الأصيل أو الواقعى أو الحقيقى Authentic Assessment، والتقويم المعتمد على الأداء Performance Assessment.

ويعرفة (صلاح الدين علام، ٢٠٠٤، ١٠٥) بأنة إظهار المتعلم بوضوح، أو يبرهن، أو يقدم أمثلة، أو تجارب، أو نتاجات أو غير ذلك لتحقيق مستوى تربوياً أو هدفاً تعلىمياً معيناً، ويمكن تقييم أداء الطالب داخل الصف الدراسي من خلال: المحادثة الشفوية، والتعبير التحريري، وإجراء التجارب المعملية، وتكوين المجسمات، ورسم الخرائط، وغير ذلك. ويمكن فى تقييم الأداء تقييم العمليات المتضمنة فى الأداء أثناء تنفيذه، ويمكن أيضاً تقييم النتاجات النهائية وتقدير درجة جودتها استناداً إلى موازين تقدير، وأهم ما يميز هذة المهام أنها مباشرة، ووظيفية، وواقعية أي تماثل مواقف حياتية فعلية خارج نطاق الصف المدرسي.

ويعتبر أحد الأساليب المتنوعة أو الأدوات المستخدمة فى التقويم التربوى البديل، ويعرف بأنة أحد المقاييس التى تتطلب من الطالب أداء شئ ما مثل حل مسألة رياضية أو كتابة مقال صحفى أو تصميم ملصق أو... غيرة، وتقدير أدائة أو ملاحظتة بواسطة أدوات قياس معينة فى ظروف عمل حقيقية أو ظروف عمل محاكية لها، ومن ثم الحكم على هذا الأداء وفق معايير محددة سلفاً (يسرى زكى، ٢٠١٦، ١٨١٥).

ويعرفة (عبد العزيز محمد، ٢٠١٩، ٥٣٦) على أنة توظيف للطالب معلوماتة ومهاراتة فى مواقف حياتية حقيقية أو مواقف تحاكى المواقف الحقيقية، أو إجرائة للبحوث المتعلقة بموضوعات المقرر الدراسى، أو أدائة لبعض الاختبارات العملية، أو قيامة بأنشطة أو عروض أو أداءات عملية يظهر من خلالها مدى إتقانة لما أكتسبة من مهارات.

أسباب ظهور تقييم الأداء

سيطرت الاختبارات الموضوعية ذات الاختيار من متعدد لمدة زمنية كبيرة على الاختبارات وذلك بسبب سهولة إعدادها، وتطبيقها، وتغطيتها للمحتوي الدراسى، وسهولة تصحيحها، وتميزها بدرجة عالية من الموضوعية والصدق والثبات، إلا أنها غير قادرة على قياس مهارات التفكير

العليا كالقدرة على حل المشكلات، والتحليل، والاستدلال، ومهارات التفكير؛ فهى تهتم بقياس المعارف الفعلية والاتجاهات السلوكية للطلاب فى مجال معين (Haladyna, 2004) إلا إنها غير قادرة على إنتاج إجابات مبنية من الطلاب فهو يختار إجابة من بين مجموعة من الاجابات.

لذا فقد أشار لين وآخرون (Linn et al., 1991) إلى مجموعة من الانتقادات للاختبارات الموضوعية تتمثل فى: التأكيد على المعارف الفعلية والمهارات المنفصلة، والسرعة فى الاجابة، والاجابات القصيرة المصطنعة، وعدم القدرة على تغطية مختلف محتويات أومجالات المعرفة، والتدريس من أجل الاختبار.

لذا فرضت تلك الانتقادات على المهتمين بالتربية بإيجاد بدائل جديدة أكثر فاعلية لتلافى عيوب التقييم التقليدي، ولتحقيق ذلك اتجهت الجهود إلى الاهتمام بالتقييمات البديلة المعتمدة على الأداء لأثارة عمليات التفكير العليا. كما لم تتوقف الجهود إلى إنجاز مهام مجردة من السياق وبعيدة عن الواقع، ففى أواخر سنوات الثمانينات زاد الطلب بكثرة لاستخدام التقييم الواقعي للاداء (Keller et al., 2010).

كما يركز تقييم الأداء على قدرة تنفيذ مهام محددة بنجاح بدلاً من اختزال بعض المعلومات والمعارف المتفرقة، لذا يمكن تفسير التحصيل من خلالة بأنة أداء أو قدرة على إنجاز شئ ما، ويتضمن الأداء توسيع نطاق المعرفة والفهم وتنمية مهارات متعددة ومتكاملة يصعب قياسها وفق الأساليب الكلاسيكية (صلاح الدين علام، ٢٠٠٠).

مكونات وأقسام تقييم الأداء

يري (صلاح الدين علام، ٢٠٠٤)، و(عبداللة السعداوى، ٢٠١٠) أن تقييم الأداء يشتمل على مكونين رئيسيين هما مهام الأداء، وقواعد تقدير الأداء، وفيما يلى عرض تلك المكونات:

١) **مهام الأداء:** وتتطلب إجراء عمليات أو سلسلة من الانشطة، أو أداء عمل معين بطريقة مناسبة، أو إبتكار وتكوين نتاجات مركبة تحقق مستويات معينة من الجودة، وعلى الرغم من إتساع نطاق مهام الأداء وتنوع صيغها وأنماطها إلا أنة يمكن تصنيفها إلى:

أ) **مهام محدودة:** تتطلب أن يقوم الطالب بأداء محدد فى زمن قصير، وتستخدم هذة المهام المحددة عادة لأغراض المساءلة، والتحقق من كفايات الطالب فى مهارات معينة ذات أهمية، مثال على ذلك: كتابة مقال من خلال استدعاء عمليات التحليل والمقارنة والاستنتاج والتقويم أو غيرها من المهام.

ب) **مهام موسعة:** تتطلب أن يختار الطالب موضوعاً معيناً ويعطى وقتاً كافياً للتفكير فية،

وبحثة من جوانب متعددة لكي يظهر تمكنه من الموضوع وعمق فهمه له، مثال على ذلك: إجراء تجربة علمية تتعلق بالتمثيل الضوئى وتفسير نتائجها، أو مراجعة قصيدة شعر والتعليق عليها، أو غير ذلك.

جـ) **عرض الاعمال:** هو توظيف الطالب لمهاراتة المتنوعة لإبراز كفاءتة فى مجال معين فهو بمثابة تقويم ختامى لانجازات الطالب.

د) **ملفات الاعمال (البورتفوليو):** تشمل على عينة من الأعمال المتجمعة المتنوعة التى قام بها الطالب خلال مدة دراسية معينة، ويمكن أن تتضمن هذة الأعمال نتاجات ابتكرها الطالب، أو أفضل الأعمال التى أنجزها خلال مدة معينة، أو بعض الأعمال التى لا تزال فى مرحلة الإعداد لكي يبين تحسنة خلال المدة المحددة، وتقويمه لجوانب القوة والضعف فى بعض هذه الأعمال.

ه) **مهام أخرى:** تشمل على أى أنشطة أخرى قام بها ودونها الطالب، أو لاحظها المعلم وحكم على نوعيتها وجودتها.

كما يرى (عبد العزيز محمد، ٢٠١٩، ٥٥٠) أن الأقسام الرئيسية الخمسة لمهام الأداء أشكالاً متعددة تتطلب جميعها إنشاء أو تكوين إجابات، أو ابتكار نتاجات تظهر قدرة الطالب على استخدام وتوظيف معارفة ومهاراتة فى مجال معين، ومن تلك الاشكال:

-**الإجابات الحرة أو المستفيضة أو الكتابة:** تتمثل فى أسئلة المقال التى تتطلب أن يُكون الطالب إجابتة بنفسة بدلاً من أن يختار الإجابة الصحيحة من بين بدائل معطاه، وتسمح تلك الإجابات بتوظيف مهارات عقلية عليا كالفهم المتعمق أو الاستدلال، والتحليل، والتفكير الناقد، وحل المشكلات.

-**الاجابات المحدودة:** تتمثل فى أسئلة الاختيار من متعدد، وأسئلة الصواب والخطأ، وأسئلة المزاوجة، وغيرها من الأسئلة بشرط أن تعكس مهام واقعية من الحياة العملية وليست أجزاء مقتضبة من المعلومات أو الحقائق والمطلوب من الطالب تذكرها.

-**التعبير الشفوى:** يعد من الأنشطة الأدائية التى ينبغى أن يكتسبها الطلاب، ومن المهارات التى تتعلق بالتواصل اللغوى ويتم تقويمها شفوياً: القراءة، والتعبير، وإلقاء الكلمات، والمناظرات، والمناقشات، وتظهر قدرة الفرد على التعبير والتلخيص وربط الأفكار.

-**عروض الأعمال:** تعد من المهام الأدائية المتسعة النطاق، وتتطلب إظهار الطالب كفاياتة المتعددة فى تصميم أعمال معينة، وتتعدد صيغ عروض الأعمال مثل المقال، أو الورقة

البحثية، أو المقابلة، ـــأو العروض الدرامية، أو غيرها.

-**التجارب:** تعد التجارب من المهام الأدائية الأساسية فى تعلم الرياضيات، والعلوم بخاصة، وتتضمن هذة التجارب ممارسة الطالب الفعلية لعمليات التخطيط والتصميم وإجراء التجارب المعملية فى مواقع العمل بدلاً من أختبارات الورقة والقلم.

٢) قواعد تقدير الأداء: هى عبارة عن محكات الحكم على جودة مهام الأداء أو نوعيتة، وترتبط ارتباطاً وثيقاً بمهام الأداء حيث أن كلا منهما يعتمد على الآخر ويكمله، وتساعد تلك القواعد على تصميم مهام يمكن تقدير مكونات عملياتها أو نتاجاتها، ويمكن أن تنقسم إلى:

أ) **قواعد تقدير محدودة:** هى عبارة عن قواعد توضح خصائص الأداء الجيد للمهام المحدودة وما يشتمل علىها من مهارات وكفايات، وهذة القواعد تكون نوعية بمعني أقتصارها على مهام محددة.

ب) **قواعد تقدير عامة:** هي عبارة عن قواعد توضح خصائص الأداء الجيد للمهام الموسعة وما يشتمل علىها من مهارات وكفايات، ويمكن أن تنطبق هذة القواعد على مهام متنوعة ومختلفة.

جـــ) ملاحظات المعلم المنظمة: من خلال ملاحظة أداء الطلاب وسلوكهم، ويمكن أن تكون هذه الملاحظات موضوعية من خلال ربطها بمستويات المحتوي، وتحليل مكونات الأداء المراد ملاحظتة فى مواقف طبيعية، وإعداد قواعد أو محكات لتقدير هذة المكونات والحكم على نوعيتها.

خصائص مهام الأداء

إن مهام الأداء تتميز بعدة خصائص تميزها عن التقويم التربوى التقليدى، ومن تلك الخصائص:

١- **الارتباط بالمستويات أو النواتج التعليمية:** فالتقويم التربوى البديل ومهام الأداء تهدف إلى تحقيق مستويات تربوية متميزة أو نواتج تعلىمية محددة تتطلب إبراز مهارات الطالب المتنوعة التى تتميز بالواقعية، فى إطار مناهج دراسية قائمة على التفكير وعمليات تدريس تؤكد على الانشطة التعليمية وشمولها وإيجابية الطالب ومشاركتة بفاعلىة وتوجيهه ومساعدتة على بناء معرفتة من خلال تفاعلة مع بيئته ومرورة بخبرات متنوعة بدلاً من عمليات التدريس القائمة على التلقين من جانب المعلم (محمد بن راشد، ٢٠٢١، ١٤٧).

٢- **الوضوح:** بمعني وضوح المطلوب أداؤة أو تكوينة، وشروط الأداء من حيث حدود الوقت،

والمصادر المتاحة، والاستعانة بالآخرين، ومحكات تقدير الأداء والحكم عليه، وصياغة كل تلك التعليمات فى عبارات بسيطة لضمان فهم جميع الطلاب لها (عادل سرايا، ٢٠٠٥).

٣- **واقعية المهام:** يقوم تقييم الأداء على حل مشكلات واقعية أو كتابة أسلوب الحل الذى يراه مناسباً أو عرض الحل شفوياً أو بأى طريقة مماثلة تعتمد على عمليات عقليا عليا (تهانى المزينى، ٢٠١٥).

٤- **التقييم المباشر للسلوك أو الأداء:** يسعى التقويم البديل إلى تقويم أداء الطالب وسلوكة بطريقة مباشرة لذا فيجب أن تكون المهام المستخدمة فى العمليات الاختبارية تمثل عينة ممثلة للمهارات، ويقوم المعلم بالحكم على سلوك أو أداء أو نتاجات الطالب بناءاً على محكات تقدير محددة وواضحة وتقدير درجاتها يتم بواسطة مجموعة من المصححين المؤهلين ويجب أن تكون تلك المحكات المستخدمة فى التقدير واضحة أيضاً للطالب من أجل توجيهم أنشطتهم وتعلمهم(الفريق الوطنى للتقويم، ٢٠٠٤، ١١- ١٢).

٥- **الاستناد إلى عينات من الأداء عبر الزمن:** مهام الأداء المختلفة تجرى على الطلاب عبر الزمن، وذلك لانها تهتم بفحص أنماط أعمال الطالب المختلفة وخططة ومخططاته ونموة فيها لتعطى صورة نهائية كاملة عن أداء الطالب النهائى، لذا تتسم مهام الأداء بالاستمرارية (حسن زيتون، ٢٠٠٨).

٦- **شمولية وتكامل التقويم:** فتقييم الأداء شمولي لانة يعتمد على إستخدام أدوات ووسائل متنوعة لتقييم أداء الطالب فى مواقف التعلم المختلفة من مختلف جوانبة ليست المعرفية فقط بل يمتد ليشمل المهارات والمعارف واتجاهات الطلاب وعاداتهم وسماتهم (فوزية الدوسرى، ٢٠٠٥).

تقييم أداء المعلم

تقييم الأداء يمثل مدخلاً لتطوير أداء المعلم، حيث أن توظيف التقويم الحقيقى يتطلب من المعلم الانتقال من ثقافة الصف الدراسي القائم على عمليات التذكر والاسترجاع والمهام التنافسية والاختبارات ذات الفرصة الواحدة والمقارنات بين الطلاب إلى الصف القائم على المهام التعاونية وربط المادة العلمية بالمواقف الحياتية والمشكلات ذات الصلة ببيئة الطالب، وتنويع المهام، وتنويع استراتيجيات وأدوات التقويم، ومقارنة الطالب بذاتة لدراسة مدي التحسن فى أدائة (يحى عبدالخالق، ٢٠١٨، ٢٩٥).

ومما يؤكد ذلك ما أشار لة (على عبد العظيم، ٢٠١٥، ٩٤) بأن التقويم البديل يُعد الحل المناسب للحكم على أداء المعلم وتطوير برامج إعدادة، وهذا ما يؤكدة أيضاً دراسة (محمد موسى،

٢٠١٨) من وجود ارتباط موجب دال احصائياً بين درجة استخدام أدوات التقييم الواقعى وبين درجة التطور المهني الذاتي للمعلمين.

كما أن أستخدام التقويم التربوي البديل أو الواقعي فى تقييم أداء المعلم يهدف إلى تقويم الجوانب المختلفة فى شخصية المتعلم المعرفية والمهارية والوجدانية وليس الجانب المعرفى فقط (حجاج غانم، ٢٠٠٧، ٢).

وتؤكد دراسة (محمد يوسف، ٢٠١٧) أن التقويم البديل هو أساس عملية إعداد الطالب المعلم؛ وذلك لانة يتضمن العوامل التى تمكن الطالب من مواكبه متطلبات العصر وإكتساب المهارات اللازمة التى تؤهلة لأداء مُرضي، لذا فقد أوصت هذة الدراسة بضرورة أستحداث أساليب ووسائل الإعداد والتدريب فى كليات التربية المختلفة، وإستخدام التقويم البديل كأحد حلول عملية الاعداد بما يسهم فى إعداد الطالب المعلم وتنمية أفكاره والخروج من قيود الاعداد التقليدى الذي يبدأ بالتدريس الالقائي وينتهى بالاختبارات ومنح الدرجات إلى فكر ووسائل وأدوات التقويم التربوي البديل.

ويري (على عبد العظيم، ٢٠١٥، ٩٤) إن التدريس عملية تتطلب مهارات وقدرات وسمات شخصية ومعرفة بمجال التخصص تتضامن جميعها لرسم صورة كاملة عن أداء المعلم؛ وقد يؤدي قياس كل جانب منها منعزلاً عن الجروانب الآخري خارج نطاق الموقف التعلىمي إلى نتائج مضللة عن الأداء، ويعد التقويم البديل الحل المناسب للحكم على أداء المعلم وتطوير برامج إعدادة.

لذا يرى (صفوت توفيق، ٢٠٢١) بضرورة إعتماد استراتيجيات وأدوات ومفاهيم التقويم التربوى البديل ضمن مقررات برامج إعداد المعلمين فى كليات التربية المختلفة.

منهج البحث:

أغلب الابحاث السيكومترية تعتمد فى الاصل على الطبيعة الوصفية الكشفية، لذا فأن المنهج المناسب لتلك الدراسة هو المنهج الوصفى الذى يهتم بإستكشاف مكونات التباين الأكثر تأثيراً على ثبات وصدق درجات التقييم، ومعاملات إمكانية التعميم النسبية والمطلقة، وإستكشاف الأبعاد ذات الفعالية فى زيادة إمكانية التعميم وثبات درجات أداء الطلاب للمهام المختلفة .

إجراءات البحث:

١- مجتمع البحث:

تكون مجتمع البحث من طلاب الفرقة الرابعة بكلية التربية – جامعة السويس من الشعب العلمية والأدبية للتعليم العام (إعدادي وثانوي) والتعليم الاساسي (إبتدائي) وشعب (تربية فنية

ورياض أطفال وتكنولوجيا التعليم والتربية الخاصة) في العام الجامعي ٢٠٢١ / ٢٠٢٢م، حيث بلغ عدد مجتمع الدراسة (٦٤٢) طالب وطالبة بمتوسط حسابي (٢١,٣) لاعمارهم وإنحراف معياري (٠,٦٤٣).

٢- عينة البحث الاستطلاعية

تكونت العينة الاستطلاعية من (٢٠٠) طالب وطالبة من طلاب الفرقة الرابعة بكلية التربية – جامعة السويس من الشعب العلمية والادبية للتعليم العام (إعدادى وثانوى)، بمتوسط عمر زمنى = ٢١,١، وانحراف معيارى = ٠,٦٨٧، وتم إستخدام العينة للاستطلاعية للتحقق من بعض التحليلات الأولية لاختبار المواقف بإستخدام النظرية التقليدية للتحقق من الخصائص السيكومترية لها قبل إجراء تحليلات إمكانية التعميم.

٣- عينة البحث الاساسية:

تكونت عينة الدراسة النهائية من (٢٧١) طالب وطالبة من طلاب الفرقة الرابعة بكلية التربية – جامعة السويس من الشعب العلمية والادبية للتعليم العام (إعدادي وثانوي) و التعليم الاساسي (إبتدائي)، بمتوسط عمر زمني = ٢١,٦، وانحراف معياري = ٠,٧٣٤،

ثانياً: أداة البحث:

أختبار المواقف (إعداد فوزى عزت علي وآخرون، ١٩٩٧و تعديل الباحث، ٢٠٢٢)

أعد أختبار المواقف فريق عمل من المركز القومي للامتحانات والتقويم التربوي- فوزى عزت على وآخرون (١٩٩٧)، كما قام فوزى عزت (١٩٩٨) بالتحقق من البنية العاملية لأختبار المواقف بإعتبارة أختباراً للاستعداد للقبول بكليات التربية حيث قام بتحليل الدراسات السابقة و بعض المقاييس و إستمارات تحليل العمل لمجموعة مختلفة من المهن بالاضافة إلى تحليل مهنة التدريس بشكل خاص وعرض قائمة بذلك علي مجموعة من الخبراء، وقد تم تحديد خمسة متغيرات رئيسية وهامة يمكن أن تسهم في النجاح التعليمي والمهني بمهنة التدريس وهذة المتغيرات تتمثل فى: الذكاء الاجتماعى، والمبادأة، وروح الدعابة والفكاهة، وإتخاذ القرار، والاتجاة نحة مهنة التدريس.

وقام فريق العمل بإعداد مواقف تقيس هذة الابعاد وتم عرضها على بعض الخبراء من أساتذة علم النفس في مجال القياس النفسي للتأكد من مدى صحتها والوصول إلى الصورة النهائية للاختبار، وتكونت الصورة النهائية للاختبار من خمسة مواقف لكل بُعد، وكل موقف لة أربعة بدائل منها بديل واحد فقط هو الاجابة الصحيحة، وإذا قام المفحوص بإختيارة يحصل على درجة

واحدة فقط ليصبح المجموع الكلى لدرجات الاختبار هو خمسة وعشرون درجة.

وقام فوزى عزت (١٩٩٨) بحساب ثبات الاختبار بإستخدام معامل ثبات ألفا كرونباخ لأبعاد الأختبار والدرجة الكلية للاختبار، وتراوحت معادلات الثبات بين ٠,٦: ٠,٧ وهى معاملات دالة إحصائياً أى أن الاختبار لة درجات عالية من الثبات.

وقام الباحث (٢٠٢٢) بإجراء بعض التعديلات على أختبار المواقف تمثلت فى:

١- إضافة بُعدين آخرين على أبعاد الاختبار وهما (الرغبة في التعلم، الميل لمعرفة النمو الانساني والتعلم) بناءاً على تحليل عمل مهنة المعلم، وتم تمثيل كل بعد من الابعاد السابقة بعدد خمسة مواقف ولكل موقف أربعة بدائل مناسبة لة تطابقاً لاختبار المواقف الاصلى.

٢- تم تعديل طريقة تصحيح اختبار المواقف بدلاً من وجود إجابة واحدة صحيحة يحصل عليها المفحوص علي درجة واحدة فقط، فقد لوحظ تدرج البدائل وصلاحيتهم أن يكونوا إجابات للموقف متدرجة. لذا قام الباحث بتدريج البدائل من الأقوي للأضعف لتصبح (٤، ٣،٢، ١) وبالتالى لا توجد إجابة صحيحة وآخرى خاطئة ولكن تتدرج الاجابات على الاختبار ككل.

عرضت الصورة الاولية لأختبار المواقف بعد أضافة البعدين علي عينة من المحكمين والخبراء من أساتذة التربية وعلم النفس التربوي والقياس والتقويم، وبناءاً علي رآى الخبراء تكون أختبار المواقف بصورتة النهائية بعد التحكيم من سبعة أبعاد مختلفة، ويمثل كل بُعد خمسة مواقف لكل موقف أربعة بدائل متدرجة ليصبح أكبر درجة يمكن أن يحصل عليها المفحوص فى الاختبار ككل هي مائة وأربعون درجة وأقل درجة يمكن أن يحصل عليها هي خمسة وثلاثون درجة، ولا يوجد للاختبار زمن محدد، و يجيب عنه الفرد دون الحاجة إلى أرشادات أو توجيهات، وفى العادة يستطيع المجيب الإجابة على أختبار المواقف بصورتة النهائية فى فترة زمنية من ٢٥ دقيقة إلى ٣٥ دقيقة.

- **صدق الاختبار** قام الباحث بإخضاع أبعاد أختبار المواقف للتحليل العاملى الاستكشافى المحدد العوامل للتعرف على تشبعات الأبعاد التى يتكون منها الاختبار، حيث تم حساب الصدق العاملى بطريقة المكونات الأساسية ثم طريقة بروماكس وليس التدوير المتعامد لفاريماكس بسبب عدم إستقلالية الأبعاد المستخدمة على العينة الاستطلاعية التى قوامها ٢٠٠ طالب وطالبة، وتمَّ تحديد عدد العوامل لعامل واحد للتحقق من بنية الاختبار، والجدول التالى يوضح تشبعات الابعاد على العامل المحدد الذى بلغ نسبة التباين الكلية المفسرة للدرجات لة (٤٢,٣٤٠%) من التباين الكلى:

جدول (١) يوضح تشبعات أبعاد أختبار المواقف على عامل واحد

البعد	التشبع
الذكاء الاجتماعي	٠,٥٤٧
المبادأة	٠,٤٦٧
روح الدعابة والفكاهه	٠,٣٧٩
اتخاذ القرار	٠,٦٠٩
الاتجاة نحو مهنة التدريس	٠,٣٥٩
الرغبة في التعلم	٠,٦٣٢
الميل لمعرفة النمو الانساني والتعلم	٠,٤٨٩

ويتضح من الجدول السابق أنَّ ما توصل إلية التحليل العاملى يتفق مع الأبعاد التى أفترضها الباحث مما يدل علي أن المقياس يتصف بصدق بنائى مرتفع، ويدل على ترابط بين الأبعاد المكونة للعامل وتجانسها، وبالتالى تم التحقق من صدق التكوين الفرضى لأختبار المواقف.

– **ثبات الاختبار** كما قام الباحث بعد الانتهاء من التحليل العاملى الاستكشافى بحساب ثبات الأختبار بأستخدام طريقة إعادة تطبيق الاختبار بفاصل زمني قدرة (١٥ يوم) علي العينة الاستطلاعية التي بلغت (٢٠٠) طالب وطالبة، كما تمَّ أستخدام طريقة ماكدونالد أوميجا بدلاً من ألفا – كرونباخ لان حساب معامل ثبات ألفا كرونباخ يفترض تحقق بعض الافتراضات مثل: أفتراض أن تقيس جميع المفردات نفس البنية بنفس درجة الدقة، وإفتراض التوزيعات الطبيعية المستمرة للمفردات، وإفتراض التوزيع الطبيعي للنتائج الاجمالية، وإفتراض عدم أرتباط أخطاء المفردات، وإفتراض أحادية البعد، وإفتراض البيانات المتصلة (Yang& Green, 2011) وهو ما يصعب تحقيقة بشكل كامل فى أداة القياس وبيانات العينة الاستطلاعية، والجدول التالى يوضح قيم معاملات الثبات:

جدول (٢) يوضح ثبات أختبار المواقف

البعد	إعادة التطبيق	معامل ثبات أوميجا ω
الذكاء الاجتماعى	٠,٧٣٢	٠,٧٤٧
المبادأة	٠,٧١٤	٠,٧٣٧
روح الدعابة والفكاهه	٠,٧٨٤	٠,٧٩٨
اتخاذ القرار	٠,٧٦٠	٠,٧٩٩
الاتجاة نحو مهنة التدريس	٠,٧٤٦	٠,٧٥٨
الرغبة في التعلم	٠,٧٤٢	٠,٧٦٨
الميل لمعرفة النمو الانساني والتعلم	٠,٧٧٧	٠,٧٩١
أختبار المواقف ككل	٠,٧٤٢	٠,٧٩٨

ويتضح من خلال الجدول السابق أنَّ أختبار المواقف لتقييم أداء الطالب المعلم يتصف بثبات عالى على إختلاف الطرق المستخدمة فى حساب الثبات، مما يدعو إلى الثقة فى النتائج التي يمكن التوصل إليها عند أستخدام الاختبار وصلاحيتة للتطبيق.

– **تصميمات الدراسة لإمكانية التعميم:** تعتمد تقييمات الاداء على إعداد اختبارات متنوعة ومتعددة تسمح بتغطية أغلب جوانب المعلم الشخصية والنفسية والعقلية كما تحتاج إلي عدد من المصححين لتصحيح منتوج أداة تقييم الاداء بالاضافة إلى معاينة الطالب المعلم خلال فترات زمنية مختلفة حيث أن إصدار أحكام لا يكفى خلال فترة زمنية واحدة فقط، وتؤدى هذة الابعاد إلى إحداث تباينات فى الأداء لذا سوف نعتمد في تلك الدراسة على تصميم يأخذ بعين الاعتبار مختلف أبعاد موقف القياس التى يمكن أن تؤثر على درجات الطالب المعلم وهى: الأبعاد، المصححين، الفترات.

١– تصميمات الملاحظة والقياس: أشتمل تصميم الملاحظة على أربعة أبعاد متقاطعة فى موقف القياس وهما "طالب×بُعد×مصحح×فترة"، وأُعتبر الطلاب موضوعاً للقياس(بعداً للتمييز)، وأعتبرت باقى المصادر وهى الابعاد والمصححين والفترات مصادر للتباين (أبعاد الاداتية).

جدول (٣) تصميم الملاحظة ثلاثى البُعد لأختبار المواقف لتقييم أداء الطالب المعلم

الابعاد	رمز البعد	تصنيف البعد	المستويات الملاحظة	عدد المستويات
الطلاب	P	بعداً للتمييز	من الطالب ١ إلى الطالب ٢٧١	٢٧١
الابعاد	T	أبعاداً للاداتية	أختبارات المواقف من بعد ١ إلى بعد ٧	٧
المصححين	R		من المصحح ١ إلى المصحح ٢	٢
الفترات	O		من الفترة ١ إلى الفترة ٢	٢

٢– تصميمات دراسات القرار: تهتم هذة المرحلة بمحاولة زيادة معاملات إمكانية التعميم من خلال زيادة أو خفض أو حذف أو إضافة أبعاد لموقف القياس سواء بزيادة عدد الاختبارات أو زيادة عدد المصححين أو زيادة عدد الفترات أو خفضهم أو حذف بُعد أو إضافة بُعد ، وتسمى تلك المرحلة أيضاً بمرحلة التحسين.

ثالثاً: الاساليب الاحصائية:

١– معامل ثبات أوميجا، والارتباط والتحليل العاملى الاستكشافى بأستخدام برنامج SPSS

٢– معاملات إمكانية التعميم النسبية والمطلقة ومكونات التباين المختلفة بأستخدام برنامج EduG

رابعاً: عرض نتائج البحث ومناقشتها:

١) عرض نتائج دراسة إمكانية التعميم

جدول (٤)

تحليل التباين للتصميم الثلاثى البعد "طالب×بُعد×مُصحح×فترة" لأختبار المواقف

			المكونات					
مصدر التغير	مجموع المربعات	درجات الحرية	متوسط المربعات	العشوائية	المختلطة	المُصححة	النسبة (%)	الخطأ المعياري
طالب (P)	٢٤٠٩٥,٠٤٩	٢٧٠	٨٩,٢٤٠	٢,٦٨٢٥	٢,٦٨٢٥	٢,٦٨٢٥	٣٨,١	٠,٢٧٤٠
بُعد (T)	٧١٦,٥٠٠	٦	١١٩,٤١٦	٠,٠٩٦٠	٠,٠٩٦٠	٠,٠٩٦٠	١,٤	٠,٠٥٥١
مُصحح (R)	٥,٩٧٩٠	١	٥,٩٧٩٠	٠,٠٠١٨	٠,٠٠١٨	٠,٠٠١٨	٠,٠	٠,٠٠١٣
فترة (O)	١٦,٤٢١٩	١	١٦,٤٢١٩	٠,٠٠٣٢	٠,٠٠٣٢	٠,٠٠٣٢	٠,٠	٠,٠٠٣٦
طالب- بُعد (P×T)	٢١٩٩٤,٧١٤	١٦٢٠	١٣,٥٧٧٠	٢,٨٦٥٢	٢,٨٦٥٢	٢,٨٦٥٢	٤٠,٧	٠,١٢٠٨
طالب- مُصحح (P×R)	١٣٢,٩٨٥	٢٧٠	٠,٤٩٢٥	٠,٠٠٤٩	٠,٠٠٤٩	٠,٠٠٤٩	٠,١	٠,٠٠٤٥
طالب-فترة (P×O)	٧١٢,٨٢٨	٢٧٠	٢,٦٤٠١	٠,٠٣٨٨	٠,٠٣٨٨	٠,٠٣٨٨	٠,٦	٠,٠١٧٣
بُعد - مُصحح (T×R)	١,٦٠٤٩	٦	٠,٢٦٧٥	٠,٠٠٢٠	٠,٠٠٢٠	٠,٠٠٢٠	٠,٠	٠,٠٠١٣
بُعد -فترة (T×O)	٣٠,٠٨٠٩	٦	٥,٠١٣٥	٠,٠٠٣٧	٠,٠٠٣٧	٠,٠٠٣٧	٠,١	٠,٠٠٤٨
مُصحح-فترة (R×O)	٠,٠١٠٧	١	٠,٠١٠٧	٠,٠٠٠٧	٠,٠٠٠٧	٠,٠٠٠٧	٠,٠	٠,٠٠٠٤
طالب- بُعد - مُصحح (P×T×R)	٧٨٣,١٨٠٨	١٦٢٠	٠,٤٨٣٤	٠,٠١٩٨	٠,٠١٩٨	٠,٠١٩٨	٠,٠	٠,٠١٢٥
طالب- بُعد - فترة (P×T×O)	٣٤٩٢,٤١٩١	١٦٢٠	٢,١٥٥٨	٠,٨١٦٤	٠,٨١٦٤	٠,٨١٦٤	١١,٦	٠,٠٣٨٩
طالب- مُصحح- فترة (P×R×O)	١٢٥,٢٣٩٣	٢٧٠	٠,٤٦٣٨	٠,٠٠٨٥	٠,٠٠٨٥	٠,٠٠٨٥	٠,٠	٠,٠٠٦٣
بُعد - مُصحح- فترة (T×R×O)	٨,٢١٩٠	٦	١,٣٦٩٨	٠,٠٠٣١	٠,٠٠٣١	٠,٠٠٣١	٠,٠	٠,٠٠٢٥
طالب- بُعد - مُصحح- فترة (P×T×R×O, e)	٨٤٧,٢٨١٠	١٦٢٠	٠,٥٢٣٠	٠,٥٢٣٠	٠,٥٢٣٠	٠,٥٢٣٠	٠,٠	٠,٠١٨٤
المجموع	٥٢٩٦٢,٥١٣	٧٥٨٧					%١٠٠	

نلاحظ من الجدول السابق أن نسبة التباين الحقيقي (التباين المنتظم) الذى يصف الفروق الحقيقية بين الافراد فى أختبارات صيغة أختبار المواقف من بطارية تقييم أداء الطالب/المعلم بلغت (٣٨,١%) من التباين الكلى. وتشير تلك النسب إلى أنة بالأخذ في الاعتبار متوسط المُصححين والاختبارات وعدد مرات التطبيق فإن هناك فروقاً حقيقية بين الطلاب فى أختبار المواقف، ويوضح الجدول التالى مكونات التباين لكل بعد ومعاملات إمكانية التعميم النسبية والمطلقة.

جدول رقم (٥)

تحليل إمكانيةُ التعميم لتصميم القياس طالب/ بُعد مُصحح فترات (P/ TRO)

مصدر التباين	تباين التمييز	تباين الخطأ النسبي	نسبةُ تباين الخطأ النسبي	تباين الخطأ المطلق	نسبةُ تباين الخطأ المطلق
طالب (P)	٢,٦٨٢٥	.....	.....	.....	.....
بُعد (T)	.....	.....	.....	٠,٠١٣٧	٢,٦
مُصحح (R)	.....	.....	.....	٠,٠٠٠٩	٠,٢
فترة (O)	.....	.....	.....	٠,٠٠١٦	٠,٣
طالب- بُعد (PT)	.....	٠,٤٠٩٣	٨٠,٥	٠,٤٠٩٣	٧٨
طالب- مُصحح (PR)	.....	٠,٠٠٢٤	٠,٥	٠,٠٠٢٤	٠,٥
طالب- فترة (PO)	.....	٠,٠١٩٤	٣,٨	٠,٠١٩٤	٣,٧
بُعد - مُصحح (TR)	.....	.....	.....	٠,٠٠٠	٠,٠
بُعد - فترة (TO)	.....	.....	.....	٠,٠٠٠٣	٠,١
مُصحح- فترة (RO)	.....	.....	.....	٠,٠٠٠٠	٠,٠
طالب- بُعد - مُصحح (PTR)	.....	٠,٠٠٠	٠,٠	٠,٠٠٠٠	٠,٠
طالب- بُعد - فترة (PTO)	.....	٠,٠٥٨٣	١١,٥	٠,٠٥٨٣	١١,١
طالب- مُصحح- فترة (PRO)	.....	٠,٠٠٠	٠,٠	٠,٠٠٠٠	٠,٠
بُعد - مُصحح- فترة (TRO)	.....	.....	.....	٠,٠٠٠١	٠,٠
طالب- بُعد - مُصحح - فترة (PTRO, e)	.....	٠,٠١٨٧	٣,٧	٠,٠١٨٧	٣,٦
مجموع التباينات	٢,٦٨٢٥	٠,٥٠٨٢	%١٠٠	٠,٥٢٤٨	%١٠٠
الانحراف المعياري	١,٦٣٧٨				
الخطأ المعياري النسبي	٠,٧١٢٨				
الخطأ المعياري المطلق	٠,٧٢٤٤				
معامل إمكانيةُ التعميم النسبي	٠,٨٤				
معامل إمكانيةُ التعميم المطلق	٠,٨٤				

يتضح من الجدول السابق أن أكبر مصدر تباين فى القياس النسبى يرجع إلى تفاعل طالب- بُعد والذى بلغت قيمتة (٨٠,٥%) وهو يوضح الاختلاف بين متوسط أداء الطلاب من بُعد إلى آخر فى أختبار المواقف، وثانى أكبر مصدر لتباين الخطأ يرجع إلى تفاعل طالب- بُعد - فترة والذى بلغت قيمتة (١١,٥%) والذى يفسر تباين ترتيب الطلاب بين أبعاد أختبار المواقف والفترات المختلفة. فى حين جاءت باقي مكونات التباين الاخرى منخفضة جداً لا تؤثر على إمكانية التعميم فقد تراوحت بين (٠,٠% و ٣,٨%).

كما يتضح من الجدول السابق أن أكبر مصدر تباين فى القياس المطلق يرجع إلى تفاعل طالب- بُعد والذى بلغت قيمتة (٧٨%). ثم يأتى مكون التباين الناتج من تفاعل طالب- بُعد - فترة في المرتبة الثانية من مصادر تباين الاخطاء حيث بلغ قيمته (١١,١%) وهو يوضح الاختلاف بين متوسط أداء الطلاب من بُعد إلى آخر فى اختبار المواقف ومن فترة إلى آخرى. بينما جاءت باقي مكونات التباين الآخرى منخفضة ولا تؤثر على إمكانية التعميم فقد تراوحت بين (٠,٠% و ٣,٧%).

كما كانت معاملات إمكانية التعميم جيدة حيث بلغ معامل إمكانية التعميم النسبي (٠,٨٤) وهي قيمة جيدة، وبلغ معامل إمكانية التعميم المطلق (٠,٨٤) وهي قيمة جيدة أيضاً مما يدل على أن درجات الطلاب في أختبارات المواقف تتسم بالثبات أو إمكانية التعميم.

وتتفق تلك النتائج مع ما توصلت إلية نتائج دراسات تقييم الأداء مثل دراسة شافيلسون وآخرون (Shavelson et al., 1993)، ودراسة ويب وآخرون (Webb et al., 2000)، ودراسة سميث وكوليكوش (Smith& Kulikowich, 2004) خاصة فيما يتعلق بمكونى تباين تفاعل طالب- بُعد/أختبار، وتباين تفاعل طالب- بُعد/أختبار - فترة، وتتفق أيضاً مع دراسة ني وآخرون (Nie et al., 2007)، ودراسة جبريل (Gebril, 2009)، ودراسة جولر وجيليال (Guler& Gelbal, 2010)، ودراسة تيلور وباستور (Taylor& Pastor, 2013)، دراسة (صبرى محمود، ٢٠١٦)، ودراسة (طباع فاروق، ٢٠٢٠) الذين أوضحوا أرتفاع كبير بمكون تباين تفاعل طالب-أختبار/بُعد مقارنة بباقي مكونات التباين الاخرى التى جاءت ضعيفة.

وتختلف نتائج الدراسة الحالية مع ما أوضحتة دراسة شين وآخرون (Chen et al., 2007)، ودراسة لي وكانتور (Lee& Kantor, 2007)، ودراسة جيبريل (Gebril, 2009) بأن أكبر مصدر للتباين يرجع إلى تفاعل الطالب مع المهام مع المُصحح الممزوج بالاخطاء العشوائية غير المقدرة، بالإضافة إلى دراسة كازانوفا وديماس (Casanova& Demeuse, 2011) التى كشفت

أن اكبر مصدر للتباين راجع إلى المُصححين، وتفاعل المُصححين مع الاختبارات.

وفيما يتعلق بمعاملات إمكانية التعميم النسبية والمطلقة فقد جاءت مقبولة إلى جيدة وتتفق تلك النتائج مع بعض الدراسات التى توصلت إلى بلوغ معاملات إمكانية تعميم تصل للحد الأدني المطلوب مثل دراسة كلا من ويب وآخرون (Webb et al., 2000)، ودراسة سميث وكوليكوش (Smith& Kulikowich, 2004) ودراسة لي وكانتور (Lee& Kantor, 2007)، ودراسة جولر وجيليال (Guler& Gelbal, 2010)، ودراسة (صبرى محمود، ٢٠١٦).

فى حين لا تتفق مع نتائج دراسة ميسبي وبارينز (McBee& Barens, 1998)، ودراسة شيفيلسون وآخرون (Shavelson et al., 1999)، ودراسة ني وآخرون (Nie et al., 2007)، ودراسة باستور (Pastor, 2013)، ودراسة (طباع فاروق، ٢٠٢٠) التى أظهرت ضعفاً فى معاملات إمكانية التعميم النسبية والمطلقة بالرغم من أقترابها للحد الأدنى المقبول وهو (٠,٨٠) كما ذكر (Bain& Pini, 1996).

ترجع منطقية زيادة مكونات تباين تفاعل طالب– بُعد إلى وجود اختلاف فى متوسط أداء الطلاب من بُعد لآخر داخل أختبار المواقف لتقييم الاداء، وذلك بسبب: إختلاف كل بُعد عن الآخر فى الطبيعة وفى التكوين وتباينها من حيث خصائصها، ودرجة تركيبها، وقد يرجع أيضا أرتفاع مكون تباين تفاعل طالب– بُعد إلى تعدد وتنوع الأبعاد داخل الاختبار حيث يشتمل أختبار المواقف على مجموعة من السمات والقدرات والميول المهنية المتحصل عليها من تحليل عمل مهنة المعلم.

كما يمكن تفسير ذلك المكون إلى وجود مشكلة معرفية ناتجة عن أنتقال آثر التعلم لدى الطلاب من أختبارات معينة إلى آخرى كما ذكرت دراسة باركيز (Parkes, 2001)، ودراسة سكالون (Scallon, 2004)، ودراسة هوانج (Huang, 2009) وذلك بسبب عدم قدرة الطلاب على نقل معارفهم من سياق الدراسة داخل الجامعة إلى السياق التطبيقى والعملى.

وفيما يخص أرتفاع مصدر تباين تفاعل طالب– بُعد – فترة، وتفاعل طالب–فترة فهذة النتيجة قد ترجع بشكل كبير إلى تغير فى أداء الطلاب وترتيبهم عبر الأبعاد المختلفة داخل الاختبار وخلال عدد مرات التطبيق بحيث أن الطلاب انجزوا بعض الأبعاد بشكل أفضل فى الفترة الثانية ولكن أنجزوا بعض الأبعاد الآخرى بشكل أقل فى الفترة الاولى؛ مما يدل على تغير الطلاب فى استراتيجيات حل الأبعاد خلال مرات التطبيق ومحاولة تحسين أدائهم على الاختبارات فى المرة الثانية.

٢) عرض نتائج دراسة القرار

وفيما يتعلق بدراسات القرار التى تتناول زيادة عدد مستويات الأبعاد لزيادة معاملات إمكانية التعميم النسبية والمطلقة، وذلك من خلال تجريب أفضل مستويات الأبعاد التي تزيد من نطاق الدرجة الشاملة وخفض تباينات الخطأ للحصول على درجات قابلة للتعميم على مواقف قياس آخرى. ولتحقيق ذلك سوف نقدم دراستين للقرار أهتمت الأولى بزيادة عدد الأبعاد وعدد المُصححين بينما أهتمت الدراسة الثانية من دراسات القرار بزيادة عدد مرات التطبيق.

- **الدراسة الأولى من دراسات القرار** والتى أهتمت بزيادة عدد الأبعاد وعدد المُصححين:

جدول رقم (٦)

دراسات القرار للتصميم الثلاثى البعد البعد "طالب× بُعد × مُصحح× عدد مرات التطبيق "

أختبار المواقف									
عدد الأبعاد	١٠	١٠	١٠	١٢	١٢	١٢	١٤	١٤	١٤
عدد المُصححين	٢	٣	٤	٢	٣	٤	٢	٣	٤
عدد مرات التطبيق	٢	٢	٢	٢	٢	٢	٢	٢	٢
إمكانية التعميم النسبي	٠,٨٩	٠,٨٩	٠,٨٩	٠,٩٠	٠,٩٠	٠,٩٠	٠,٩١	٠,٩١	٠,٩١
إمكانية التعميم المطلق	٠,٨٩	٠,٨٩	٠,٨٩	٠,٩٠	٠,٩٠	٠,٩٠	٠,٩١	٠,٩١	٠,٩١

يتضح من الجدول السابق الذى يحدد دراسات القرار بواسطة زيادة عدد الأبعاد وعدد المُصححين لأختبار المواقف أن زيادة عدد الأبعاد أفضل من زيادة عدد المُصححين في زيادة معاملات إمكانية التعميم النسبية والمطلقة، حيث أنة يتطلب (١٠) أبعاد للوصول لمعامل إمكانية التعميم النسبى (٠,٨٩) ومعامل إمكانية التعميم المطلق (٠,٨٩)، وفى حالة زيادة عدد الأبعاد إلى (١٢) بُعد يرتفع معامل إمكانية التعميم النسبى بمقدار ضئيل جداً إلى (٠,٩٠) ومعامل إمكانية التعميم المطلق إلى (٠,٩٠)، وفى حالة أرتفاع عدد الأبعاد إلى (١٤) بُعد يرتفع معامل إمكانية التعميم النسبى بمقدار ضئيل جداً إلى (٠,٩١) ومعامل إمكانية التعميم المطلق يصل إلى (٠,٩١) كما أن زيادة عدد المُصححين من (٢) إلى (٤) مُصححين لا يؤثر على زيادة معاملات إمكانية التعميم النسبية والمطلقة.

- **الدراسة الثانية من دراسات القرار** والتى أهتمت بزيادة عدد الفترات وكانت نتائجها كالتالى:

جدول رقم (٧)

دراسات القرار للتصميم الثلاثى البعد "طالب× بُعد ×مُصحح× عدد مرات التطبيق"

أختبار المواقف									
عدد الأبعاد	١٠	١٠	١٠	١٢	١٢	١٢	١٤	١٤	١٤
عدد المُصححين	٢	٢	٢	٢	٢	٢	٢	٢	٢
عدد مرات التطبيق	١	٣	٤	١	٣	٤	١	٣	٤
إمكانية التعميم النسبي	٠,٨٦	٠,٩٠	٠,٩٠	٠,٨٨	٠,٩١	٠,٩١	٠,٨٩	٠,٩٢	٠,٩٢
إمكانية التعميم المطلق	٠,٨٦	٠,٩٠	٠,٩٠	٠,٨٨	٠,٩٠	٠,٩١	٠,٨٩	٠,٩١	٠,٩٢

يتضح من الجدول السابق الذى يحدد دراسات القرار بواسطة زيادة عدد مرات التطبيق لأختبار المواقف أن زيادة عدد مرات التطبيق يؤدى إلى زيادة معاملات إمكانية التعميم النسبية والمطلقة مع زيادة عدد الأبعاد، حيث أن زيادة عدد مرات التطبيق يزيد من معاملات إمكانية التعميم مع زيادة عدد الأبعاد، فزيادة عدد الفترات من (١) فترة إلى (٤) فترات يزيد من معامل إمكانية التعميم من (٠,٨٦) إلى (٠,٩٠) في حالة وجود (١٠) أختبارات، ويزيد معامل إمكانية التعميم المطلق من (٠,٨٦) إلى (٠,٩٠)، و زيادة عدد الفترات من (١) فترة إلى (٤) فترات يزيد من معامل إمكانية التعميم من (٠,٨٨) إلى (٠,٩١) في حالة وجود (١٢) أختبار، ويزيد معامل إمكانية التعميم المطلق من (٠,٨٨) إلى (٠,٩١)، وفى حالة وجود (١٤) أختبار وزيادة عدد الفترات من (١) فترة إلى (٤) فترات يزيد من معامل إمكانية التعميم من (٠,٨٩) إلى (٠,٩٢)، ويزيد معامل إمكانية التعميم المطلق من (٠,٨٩) إلى (٠,٩٢).

وبالتالى توصلت نتائج دراسات القرار إلى أن زيادة عدد الاختبارات/الابعاد وزيادة عدد مرات التطبيق تؤدى إلي زيادة معاملات إمكانية التعميم النسبية والمطلقة أفضل من زيادة عدد المُصححين.

وتتفق تلك النتائج مع عدد من الدراسات السابقة التى تؤكد أن زيادة عدد الاختبارات وزيادة عدد الفترات تساهم بشكل كبير فى زيادة معاملات إمكانية التعميم النسبية والمطلقة، وخفض مكونات التباين التى تؤثر على التباين الكلي لدرجات أداء الطلاب مثل دراسة شافيلسون آخرون (Shavelson et al., 1993)، ودراسة رولز-بريمو وآخرون (Rulz-Primo et al.,

1993)، ودراسة شافيلسون وآخرون (Shavelson et al., 1999)، ودراسة ويب وآخرون (Webb et al., 2000)، ودراسة نى وآخرون (Nie et al., 2007)، ودراسة لى وكانتور (Lee& Kantor, 2007)، ودراسة هوانج (Huang, 2009)، ودراسة (طباع فاروق، ٢٠١٦)، ودراسة تيلور وباستور (Taylor& Pastor, 2013)، ودراسة (طباع فاروق، ٢٠٢٠)، كما أوضحت دراسة جبريل (Gebril, 2009) أن زيادة عدد المُصححين لا يؤدى إلى زيادة معاملات إمكانية التعميم، وأوضحت دراسة جولر وجيلبال (Guler& Gelbal, 2010) أن زيادة عدد الاختبارات أفضل من زيادة عدد المصححين لزيادة معاملات إمكانية التعميم النسبية والمطلقة.

وفى تحليل مسحى أجراه هوانج (Huang, 2009) من خلال مراجعة الدراسات التى أستخدمت نظرية إمكانية التعميم بين سنتى (١٩٨٠إلي ٢٠٠٦) أوضح أنة لكى نخفض من تأثير تباين الاختبارات المؤثرة على التباين الكلى لابد من إدماج أكبر عدد من الابعاد والاختبارات فى موقف القياس، وإستخدام تصميمات متقاطعة بدلاً من تصميمات متداخلة، وإدماج الفترة كبعد من أبعاد القياس، كما أتفقت النتائج التى تم الحصول عليها من الدراسة الحالية مع العديد من الدراسات مثل دراسة لوينج وآخرون (Lueng et al., 2010)، ودراسة كارمنلا (Carminla, 2010)، ودراسة كريست (Christ, 2012)، ودراسة بروجر وآخرون (Broger et al., 2012) الذين أظهروا أن تطبيق المقياس لمرة واحدة فقط يؤدى إلي أنخفاض معاملات إمكانية التعميم.

وتختلف تلك النتائج مع ما توصل إلية دراسة كازانوفا وديميوس (Casanova& Demeuse, 2011) التي هدفت إلى فحص مصادر الخطأ المؤثرة على ثبات أختبار التعبير الكتابى في اللغة الفرنسية بأستخدام نظرية إمكانية التعميم ونموذج راش متعدد الابعاد، حيث طبق الاختبار على (٣٣) طالباً، وأوضحت الدراسة أن زيادة عدد المُصححين تزيد من معامل إمكانية التعميم المطلق أفضل من زيادة عدد الاختبارات.

وتختلف جزئياً مع دراسة (عصام الدسوقى، ٢٠٠٥) التى أوضحت أن زيادة عدد المُصححين وعدد المفردات تؤدى إلى زيادة معاملات إمكانية التعميم أفضل من زيادة عدد الفترات، ولكن تتأثر معاملات إمكانية التعميم بزيادة عدد البنود مقارنة بزيادة عدد المُصححين، كما أوضحت دراسة (صبرى محمود، ٢٠١٦) أن زيادة عدد المقيمين إلى ثلاث مقيمين وزيادة عدد مرات التطبيق إلى ثلاث مرات يزيد من معاملات إمكانية التعميم للمقياس.

وترجع منطقية ذلك إلى إنة عند الأجابة على التساؤل الأول أوضحت النتائج أن مكونات التباين الاكثر تأثيراً على ثبات درجات أداء الطلاب ترجع إلى مكون تباين تفاعل طالب– أختبار،

وتباين الاختبار، وتباين تفاعل طالب– أختبار– فترة، وبدرجة أقل تفاعل طالب– فترة ولا ترجع إلى المُصحح وتفاعلاتة مع الأبعاد الآخرى، ومن أجل خفض مقادير مكونات تباين طالب– أختبار، وتباين الاختبار، وتباين تفاعل طالب– أختبار– فترة، وتباين تفاعل طالب– فترة يتطلب ذلك زيادة عدد الاختبارات وعدد مرات التطبيق كطرق أكثر فاعلية من زيادة عدد المُصححين للوصول إلى معاملات إمكانية تعميم جيدة تساعد فى تقييم أداء الطالب/ المعلم إلا أن زيادة عدد مرات التطبيق تنتج عنها تكاليف وجهود إضافية قد لا تصل إليها تكاليف وجهود زيادة عدد الاختبارات.

وبالنظر إلى نتائج الدراسة الحالية، ونتائج الدراسات السابقة فان عملية زيادة عدد الاختبارات وزيادة عدد مرات التطبيق تمثل الطرق الافضل لزيادة معاملات إمكانية التعميم النسبية والمطلقة، ولكن ذلك يؤدي إلى مزيد من الجهد والوقت والتكلفة عند تقييم الاداء.

توصيات البحث:

فى ضوء ما توصلت إليه الدراسة الحالية من نتائج يمكن اقتراح التوصيات الآتية:

١) إدراج مفاهيم ومبادئ نظرية امكانية التعميم ببرامج الدراسات العليا لباحثين الماجستير والدكتوراه لإستخدامها فى تقدير دقة القياسات وإجراءات تحسين عمليات القياس.

٢) التنمية المهنية والمعرفية للمعلمين وطلاب كليات التربية وتغيير تفكيرهم للانتقال من أساليب التقويم التقليدى إلى أساليب تقييم الأداء المختلفة، ومعرفة مصطلحاتة وأدواتة بأعتبارة مجالاً مستحدثاً.

٣) أستخدام أختبارات المواقف كأختبارات قبول للطلاب بكليات التربية لضمان الحد الادنى من الأداء المطلوب.

مراجع البحث

١) تهانى عبدالرحمن المزينى (٢٠١٥). تصور مقترح للكفايات البحثية لمعلمات العلوم المرحلة المتوسطة فى ضوء أدوار معلم القرن ٢١ بمدينة الرياض، **المجلة العربية للعلوم التربوية والنفسية، ٥(٢٣)**، ٥٣٥- ٥٦٤.

٢) حجاج غانم أحمد (٢٠٠٧). المشكلات المرتبطة بتطبيق أسلوب التقويم التربوي الشامل في المدارس الابتدائية من وجهة نظر المعلمين ذوي المعارف الكافية عن هذا الاسلوب وعلاقة ذلك ببعض المتغيرات الاسمية. **مجلة البحث في التربية وعلم النفس- كلية التربية جامعة المنيا، ٢١(١)**، ١- ٧٣.

٣) حسن حسين زيتون (٢٠٠٨). **أصول التقويم والقياس التربوي "المفهومات والتطبيقات"**. الرياض: الدار الصولتية.

٤) صبري محمود عبدالفتاح أمين (٢٠١٦). ثبات تقييمات الكفاءة الاجتماعية لدي الاطفال بأستخدام النسخة العربية لمقياس الكفاءة الاجتماعية المعدل في ضوء نظرية إمكانية التعميم. **مجلة العلوم التربوية، (٢)**، ١٣١- ١٨١.

٥) صفوت توفيق هنداوي (٢٠٢١) . برنامج تدريبي قائم علي التفكير التأملي لتنمية مهارات استخدام اساليب التقويم اللغوي البديل لدي معلمي اللغة العربية في المرحلة الاعدادية وأثرة علي تنمية مهارات الانتاج اللغوي لدي طلابهم. . **مجلة البحث العلمي في التربية- جامعة عين شمس ، ١٢(٢٢)**، ١٣٨- ١٨١.

٦) صلاح الدين محمود علام (٢٠٠٠). **القياس والتقويم التربوي والنفسي- أساسياته وتطبيقاته وتوجهاته المعاصرة.** القاهرة: دار الفكر العربي، الطبعة الاولي.

٧) صلاح الدين محمود علام (٢٠٠٤). **التقويم التربوى البديل أسسة النظرية والمنهجية وتطبيقاته الميدانية.** القاهرة: دار الفكر العربي.

٨) طباع فاروق (٢٠١٩). من النظرية الكلاسيكية للاختبارات إلي نظرية إمكانية التعميم. **جامعة مولود معمري- تيزى وزو.**

٩) طباع فاروق (٢٠٢٠). استخدام نظرية إمكانية التعميم فى تقدير ثبات أختبار تقييم كفاءة الرياضيات لدي طلاب السنة الرابعة ابتدائي. **المجلة الاردنية في العلوم التربوية. ١٦ (١)**، ١-١٨ .

١٠) عادل سرايا (٢٠٠١). التقويم الحقيقى، **مجلة التدريب والتقنية، ٤٠ (٧٤).**

١١) عبد العزيز محمد (٢٠١٩). تقويم ممارسات التقويم لدي أعضاء هيئة التدريس بجامعة المنيا في ضوء استراتيجيات التقويم البديل. **مجلة كلية التربية– جامعة أسيوط ، ٣٥ (٦)**، ٥١٩– ٥٩٦.

١٢) عبدالله السعدوى (٢٠١٠). **دليل المعلم للتقويم المعتمد على الاداء من النظرية إلى التطبيق**. الرياض: مكتبة التربية لدول الخليج العربى.

١٣) عصام الدسوقى (٢٠٠٥). استخدام نظرية إمكانية التعميم فى تقدير ثبات بطاقة ملاحظة أداء الطلاب فى التربية الميدانية. **مجلة كلية التربية بدمياط، ١(٤٨).**

١٤) علي عبد العظيم علي (٢٠١٥). التقويم البديل: مدخل للارتقاء بأداء المعلم وتطوير برامج إعدادة. **المؤتمر العلمي الرابع والعشرون للجمعية المصرية للمناهج وطرق التدريس** تحت عنوان: برامج إعداد المعلمين في الجامعات من أجل التميز، ٩١ – ١١٩.

١٥) الفريق الوطنى للتقويم (٢٠٠٤). **استراتيجيات التقويم وأدواتة (الإطار النظرى)**. المملكة الأردنية الهاشمية: وزارة التربية والتعليم.

١٦) فؤاد أبو حطب (١٩٩٢). **دليل المعلم في تقويم الطالب**. المركز القومي للامتحانات والتقويم التربوي بالاشتراك مع وزارة التربية والتعليم. القاهرة: دار غريب للطباعة.

١٧) فوزى عزت على (١٩٩٨). البنية العاملية لأختبار الاستعداد للقبول بكليا التربية. **المجلة المصرية للدراسات النفسية، ٨ (١٩)**، ٧٩– ١٠٣.

١٨) فوزى عزت على، وآخرون (١٩٩٧). **أختبار الاستعداد للقبول بكليات التربية**. القاهرة، مطبوعات المركز القومي للامتحانات والتقويم التربوى.

١٩) فوزية محمد الدوسري (٢٠٠٥). تقويم مهارات استخدام الخرائط لدى تلميذات الصف الثالث من المرحلة المتوسطة بالمملكة العربية السعودية، **مجلة الجمعية التربوية للدراسات الاجتماعية، (٤).**

٢٠) محمد بن راشد عبدالكريم (٢٠٢١) . واقع التقويم البديل في التعليم العام بمحافظة القنفذة من وجهه نظر المعلمين. **رابطة التربويين العرب. السعودية، ١٣٧(١٣٧)**، ١٤١ – ١٦٢.

٢١) محمد موسي نصر الله (٢٠١٨). واقع استخدام التقويم الواقعي وعلاقتة بالتطور المهني الذاتي لدي معلمي المرحلة الاساسية الدنيا في محافظة جنين. **رسالة ماجستير غير**

منشورة، كلية العلوم التربوية، جامعة القدس.

٢٢) محمد يوسف أحمد (٢٠١٧). التقويم التربوي البديل ودورة في تنمية كفايات الطالب/ المعلم بكليات التربية. كلية التربية جامعة بيشة. **المجلة العربية للعلوم ونشر الابحاث- المملكة العربية السعودية، ١(١).**

٢٣) ياسر عبدالحافظ و عبدالله سالم العازمي (٢٠١٣). تقويم اداء معلم التعليم الاساسي في ضوء معايير المجتمع المدرسي والمحلى. **مجلس النشر العلمي بجامعة الكويت، ٢٧(١٠٦)**، ٥٥ – ٩٩ .

٢٤) يحي عبد الخالق يوسف (٢٠١٨). المعيقات التي تواجه تطبيق التقويم الحقيقي في تعليم وتعلم مقررات التربية الاسلامية بمدارس منطقة تبوك التعليمية. **المجلة الدولية للدراسات التربوية والنفسية، ٢ (٣)**، ٢٩٢- ٣١٦.

٢٥) يسري زكي عبود (٢٠١٦). التقويم البديل كاتجاه حديث في تقويم أداء الطلاب. **المؤتمر الدولي لكلية التربية بجامعة الملك خالد** تحت عنوان: المعلم وعصر المعرفة الفرص والتحديات. كلية التربية بجامعة الملك خالد- السعودية، ١٨٠٣ – ١٨٣٨.

1) Atilla, Y. & Ezel, T. (2010). The Examination of Reliability According to Classical Test and Generalizability on a Job Performance Scale. Educational Sciences: Theory & Practice, 10 (3), 1847-1854.
2) Bain, D. & Pini, G. (1996). To rate your ratings. Generalizability: a user's guide. Geneva: psych pedagogical research center. General Management of the Orientation cycle.
3) Bakker, M., Sanders, P., Beijaard, D., Roelofs, E., Tigelaar, D., & Verloop, N. (2008). Reliability and generalizability of performance judgments based on a video portfolio. Pedagogische Studiën 85(4), 240-260
4) Bell, C. A., Gitomer, D. H., McCaffrey, D. F., Hamre, B. K., Pianta, R. C., & Qi, Y. (2012). An Argument Approach to Observation Protocol Validity. Educational Assessment, 17(2-3), 62–87.
5) Bertrand, R. & Blais, J. (2004). Models de measure. Canada: presses de L'Universite du Quebec.
6) Brennan, R. L. (2001). Generalizability Theory. New York: Springer.
7) Brennan, R. L. (2010). Generalizability Theory. Educational Measurement, 61-68. At: http://www.education.uiowa.edu.
8) Broger, T., Saintmeen, C., & Keleven, W. (2012). The generalizability of systematic rating scales across time: A preliminary analysis using generalizability theory. Journal of educational psychology,

3, 75-92.

9) Cardinet, J., Sandra, J., & Pini, G. (2010). Applying Generalizability Theory using EdUG. New York: Routledge.
10) Carminla, S. (2010). Social function assessment tools for children within the framework of the generalizability theory. Clinical psychology review, 4, 52-65.
11) Carrie, L. S. (2013). Using Generalizability theory to measure sources of variance on a special education teacher observation tool. A dissertation of doctor of education in curriculum and instruction, Boise state university.
12) Cary, Maranell.M. (1994). Scalling – Chicago: Aldine publishing company.
13) Casanova, D., & Demeuse, M. (2011). Analysis of different facets influencing the reliability of the written expression test of French as a foreign language test. Measurement and Evaluation in Education, 34(1), 25-53.
14) Christ, G. (2012). An Application of the generalizability theory to the social skills improvement system- Rating scales. North American journal of psychology, 3, 25-37.
15) Chen, E., Niemi, D., Wang, H., & Mirocha, J. (2007). Examining the generalizability of direct writing assessment tasks. University of California. Los Angeles: CRESST (CSE Technical Report N° 718). Available online at: cresst.org/publications/cresst-publication-3089.
16) Crocker, L. & Algina, J. (2006). Introduction to classical and modern test theory. Belmont, CA: Wadsworth Publishing Company.
17) Demet, H. & Erkut, A. (2018). Factors that affect the performance of teachers working in secondary-level education. Journal of Academy of Educational Leadership journal, 22(1), 1- 19.
18) Everitt, B. & Skrondal. J. (2010). The Cambridge dictionary of statistics. London: Cambridge university press.
19) Gao, X., & Brennan, R. (2001). Variability of estimated variance components and related statistics in a performance assessment. Applied Measurement in Education, 14(2), 191–203.
20) Gebril, A. (2009). Score generalizability of academic writing tasks: Does one test Methods fit it all? Language Testing, 26(4), 507-531.
21) Guler, N. & Gelbal, S. (2010). Studying Reliability of Open Ended Mathematics Items According to the Classical Test Theory and Generalizability Theory. Educational Sciences: Theory & Practice, 10 (2), 1011-1019.

22) Haladyna, T. (2004). Developing and validating multiple- choice test items. New jersey: Lawrence Erlbaum Associates publishers.
23) Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When Rater Reliability Is Not Enough Teacher Observation Systems and a Case for the Generalizability Study. Educational Researcher, 41(2), 56–64.
24) Huang, C. H. (2009). Magnitude of task-sampling variability in performance assessment: a meta-analysis. Educational and Psychological Measurement, 69 (6), 887- 912.
25) Keller, L., Clauser, B., & Swanson, D. (2010). Using multivariate generalizability theory to assess the effect of content stratification on the reliability of a performance assessment, Advanced in Health Science Education, (15), 717-733.
26) Kerry, E. (2015). Teacher Performance Appraisal: More about performance or Development. Australian Journal of Teacher Education, 40(9), pp 102- 116.
27) Lane, S., Liu, M., Ankenmann, R., & Stone, C. (1996). Generalizability and validity of a mathematics performance assessment. Journal of Educational Measurement, 33(1), 71-92.
28) Lee, Y., & Kantor, R. (2007). Evaluation prototype tasks and alternative rating schemes for a new ESL writing test through G-theory. International journal of Testing, 7(4), 353-385.
29) Linn, R., Baker, E. & Dunbar, S. (1991). Complex Performance-Based Assessment: Expectations and Validation Criteria. Educational Researcher, 20(8), 15-21.
30) Lueng, T., Lui, R. & Yim, H. (2010). Advances in the assessment of social competence: Finding from a preliminary dependability investigation in a Japanese sample. Japanese Journal of educational psychology,2, 56-79.
31) Maria, L. & Maria, H. (2013). Approaches to Teachers' Performance Assessment for Enhancing Quality of Education at Universities. Procedia - Social and Behavioral Sciences. (106), pp 476-484.
32) Mcbee, M., & Barnes, L. (1998). The generalizability of a performance assessment measuring achievement in eight-grade mathematics. Applied measurement in educations, 11(2), 179-194.
33) Meyer, J. P. (2010). Reliability. New York: Oxford University Press.
34) Mohamed, A. A. (2010). Teaching licensure as a requirement to the teacher's performance quality. مجلة كلية التربية جامعة الازهر , 144(2), pp 585- 606.
35) Moss, P. (1992). Shifting conceptions of validity in educational

measurement: implications for performance assessment. Review of Educational Research, 62(3), 229-258.

36) Nicholas, T. (2016). Generalizability of universal screening measures for behavioral and emotional risk. A Dissertation of Doctor of Philosophy, university of Arizona, ProQuest.

37) Nie, Y., Yeo, S., & Lau, S. (2007). Application of generalizability theory in the investigation of the quality of journal writing in mathematics. Studies in Educational Evaluation, 33(3-4), 371-383.

38) Parkes, J. (2001). The role of transfer in the variability of performance assessment scores. Educational Assessment. 7(2), 143- 164.

39) Ruiz-Primo, M., Baxter, G., & Shavelson, R. (1993). On the stability of performance assessment. Journal of educational measurement, 30(1), 41-53.

40) Salkined, N. J. (2008). Encyclopedia of educational psychology. Los Angeles: Sage publications.

41) Scallon, G. (2004). The evaluation of learning in a competency-based approach. Brussels: De Boeck.

42) Shavelson, R, Baxter, G., & Gao, X. (1993). Sampling variability of performance assessments. Journal of Educational Measurement, 30(3), 215- 232.

43) Shavelson, R. & Webb, N. (2009). Generalizability Theory and its contribution to the discussion of the generalizability of research finding. In K. Erichan., & W. Roth (Eds.), Generalizability from Educational Research (pp13-32). New York: Routledge.

44) Shavelson, R.J., & Webb, N.M. (1991). Generalizability theory: A primer. Newbury Park, CA: Sage.

45) Smith, E., & Kulikowich, j. (2004). An application of generalizability theory and many-facet measurement using a complex problem-solving skills. Educational and psychological Measurement, 64(4), 617- 639.

46) Suen, H.K. & Lei, P.W. (2007). Classical versus Generalizability theory of measurement. Educational Measurement, (4), 1-13.

47) Taylor, A., & Pastor, D. (2013). An application of generalizability theory to evaluate the technical quality of an alternate assessment. Applied Measurement in Education, 26(4), 279–297.

48) Urbina, S. (2004). Essentials of psychological testing. New Jersey: john Wiley& sons.

49) Webb, N., Shavelson, R. & Haertel, H. (2006). Reliability coefficient

and generalizability theory. Handbook of statistics, 26, 4-44.

50) Webb, N., Schlackman, J., & Sugrue, B. (2000). the dependability and interchangeability of assessment methods in science. Applied Measurement in Education, 13(3), 277–301.

51) Yang, Y., & Green, S. (2011). Coefficient alpha: A reliability coefficient for the 21st century?. Journal of psychoeducational Assessment, 29, 377- 392.

52) Yılmaz, N., Başbaşa,N. (2015). Assessment of sewing and picking skills station reliability with generability theory. Journal of Measurement and Evaluation in Education and Psychology, 6(1), 107-116.

53) Yılmaz, N., Gelbal, S. (2011). Comparison of different patterns with generality theory in the case of communication skills station. Hacettepe University Journal of Education Faculty, 41, 509-518.

The stability of the test of attitudes to evaluate the performance of the student teacher by generalizability theory

Prepared by/ Joseph Samir Wahib Kamel
Assistant Lecturer, Department of Educational Psychology
Faculty of Education - Suez University

The current study aims to verify the stability of the test scores of attitudes to evaluate the performance of the student teacher by generalizability theory, and to know the components of variance that most affect the stability of the scores of performance evaluation by that test, and to know the best ways to improve future measurement procedures and the most effective in measurement designs to reach generalizable scores The current study was applied to (271) male and female students from the fourth year at the Faculty of Education - Suez University, where they applied the attitude test, and the results revealed that the attitude test reached a good degree of the coefficient of relative and absolute possibility of generalization, which indicates the stability of the attitude test scores, and that the largest components The effect of variance on the coefficient of relative and absolute possibility of generalization was represented in the student-test interaction, and the student-test-period interaction. While the rest of the other variance components were weak and did not affect the generalizability coefficients, the results also showed that increasing the number of dimensions and increasing the number of application time's lead to an increase in the relative and absolute generalizability coefficients better than increasing the number of correctors.