DOI: http://dx.doi.org/10.24200/jeps.vol12iss2pp246-261

تقدير ثبات علامات الطلبة في المقررات الجامعية على الاختبارات خلال الفصل الدراسي ومدى تأثر هذه التقديرات بطريقة تحديد أوزانها النسبية

أحمد عودة* وماجد الجودة
جامعة تبوك، المملكة العربية السعودية

استُلم بتاريخ: ٢٠١٧/١/٣ قُبل بتاريخ: ٢٠١٧/١٠/١٨

ملخص: هدفت هذه الدراسة الى تقدير معاملات ثبات الاختبارات التي يعدّها أعضاء هيئة التدريس بجامعة تبوك في مقررات السنة التحضيرية في شطري الطلاب والطالبات، باستخدام معادلة راجو والنسب التي تحددها تعليمات الجامعة للاختبارات خلال الفصل، مرة، (وهي أوزان ثابتة بصرف النظر عن الاختلافات في قدراتها التمييزية أو الفروق الحقيقية بين الطلبة)، ومرة أخرى باستخدام الانحرافات المعيارية للدرجات على هذه الاختبارات، ومقارنتها بالمعاملات المحسوبة من معادلة فيلدت – جيلمر القائمة على مصفوفة التباينات والتباينات المشتركة لهذه الاختبارات. وللإجابة عن أسئلة الدراسة تم تحليل الدرجات لجميع الشعب لتلك المقررات في الشطرين، والبالغة ٦٦٣شعبة، وفي ضوء المحددات والافتراضات التي أشارت اليها الدراسة، تم تقدير الثبات بمعادلة فيلدت – جيلمر على عينة من ٣٠ شعبة مختارة عشوائيا من العدد الأصلي للشعب. وأشارت النتائج الى أن الخطة الاستراتيجية للجامعة تقوم على الجودة والتميز مما يستوجب تقديم دقة أعلى للاختبارات من القيم المتوسطة التي أشارت اليها معاملات الثبات ٠,٦٨ (عند استخدام الأوزان النسبية التي حددتها التعليمات، مقابل ٠,٦٧ ، ٠,٦٩على التوالي) عند استخدام الانحرافات المعيارية، ثم مصفوفة التباينات والتباينات المشتركة في تقدير تلك المعاملات. وكان التوافق عاليا (ارتباط = ٠,٩٨) بين تقديرات الثبات بمعادلة راجو المعدّلة وفق الانحرافات المعيارية وتلك المقدّرة وفقا للمصفوفة؛ ولذلك أوصت الدراسة باعتماد أعضاء هيئة التدريس لمعادلة راجو المعدّلة وفق الانحرافات المعيارية لتمتعها بمزايا إحصائية وعملية.

كلمات مفتاحية: معادلة راجو، الأوزان النسبية، معادلة فيلدت – جيلمر، مصفوفة التباينات والتباينات المشتركة.

Estimating the Reliability of Students' Scores in University Courses on Tests during the Semester and the Extent of Influence of these Scores Based on Determining their Relative Weights

Ahmad Audeh* & Majed Al Joudeh
Tabouk University, Saudi Arabia

Abstract: The aim of this study was to estimate the reliability coefficients of the tests prepared by faculty members in the preparatory year courses for both males and females at the University of Tabuk, using the Raju formula based on the predetermined fixed weights in the university regulations, regardless of their differentiating capability or the students' actual differences. Reliability coefficients were estimated again using the standard deviations of scores on these tests that were compared with the coefficients estimated by the Feldt-Gilmer formula that uses the variance covariance matrix of these tests. To address the questions of the study, the scores for all the students in 663 sections and 15 courses were analyzed, and the reliability coefficients were estimated. In light of the study's hypothesis, another sample of 30 sections was selected randomly from the original sample to estimate reliability coefficients through the Feldt-Gilmer formula. The results indicated that the estimated mean of coefficients 0.68 in the average were lower than expected. Based on the university strategy which focuses on quality and excellence, higher reliability is required; mediocrity in this situation is not accepted. This means that 0.68 was compared to 0.67 and 0.69 estimated by the adjusted Raju formula using standard deviations as weights, and by the Feldt Gilmer respectively. The correlation between the reliability estimations using the last two formulas was very high (r = 0.98). The results revealed enough evidence statistically and practically for recommending the use of the modified Raju formula in practical situations to estimate the tests' reliability.

Keywords: Raju Formula, relative weights, Feldt-Gilmer Formula, variances-covariance matrix.

*prof_ahmad_audeh@yahoo.com

تشير الأدبيات الحديثة أن جودة التعليم تركز بشكل كبير على مبدأ المساءلة (Accountability)، وأن قدرة الاختبارات من إعداد أعضاء هيئة التدريس على التنبؤ بتحصيل الطلبة في الاختبارات الدولية المقننة من الأساليب المستخدمة في تطبيق ذلك المبدأ (Notar, Zuelke, & Yunker, 2004) فقد أشارت دراسة نوبل Noble, (1991) ودراسة صوير (Sawer,1989) إلى أن ضعف الثبات في العلامات الجامعية كان من الأسباب وراء تدني القدرة التنبؤية لامتحانات القبول في الجامعات الأمريكية؛ وقد دفع هذا الضعف المؤسسات التعليمية الى تمكين أعضاء هيئة التدريس في الجامعات من مهارات القياس والتقويم بشكل عام، وبناء أدوات التقويم ومنها الاختبارات بشكل خاص، وذلك من خلال تقديم برامج تدريبية متخصصة لذلك الغرض، فقد كشفت الدراسات الحديثة أن خبرة المدرس، ومستوى تعليمه، وتدريبه على بناء الاختبارات، وقدرته على تحليل المحتوى، ومعرفته بمستويات وتصنيف الأهداف تؤثر بشكل كبير على درجة الصدق والثبات في الاختبارات (Iukke, 2014).

يمر بناء الاختبار وفق خطوات منظمة في ضوء المواصفات والمعايير التي أكدت عليها أدبيات القياس والتقويم؛ ولذلك تعددت مصادر الأخطاء في التقويم لتحصيل الطلبة، مما يفرض على باني الاختبار أن يتمتع بمجموعة من المهارات والكفايات المستمدة من القياس والتقويم، ويفترض الكثير من أعضاء هيئة التدريس أن مجرد تعيينهم في الجامعات يعني أنهم قادرون على التقويم بالاختبارات وبغير الاختبارات، الا أن عوامل تحقق هذا الافتراض غير متوفرة، ولذلك قد تختلف مصادر أخطاء القياس من شعبة الى أخرى، أو من اختبار الى آخر. وتشير ادبيات القياس والتقويم أيضا الى أن الاختبارات من إعداد المدرسين تعاني من مشاكل عديدة، وأنها لا تشكل مواقف تعليمية أو أدوات قياس وتقويم بالمعنى العلمي (Kingston and Nash, 2011) ولا تحقق أغراضها لأنها لا تحقق معايير الجودة ، وعلى رأسها معياري الصدق والثبات Parr and Bauer, (2006) فهما من أبرز المعايير التي تحدد الثقة بالأوزان التي تعطى للبيانات (الدرجات) التي تتحقق من اختبار أو تقرير أو مشاركة صفية كأدوات تقويم في المقرر Thorndike and (Hagen, 1986) ويعرّف الصدق بشكل عام بمدى قياس الاختبار ما أعدّ لقياسه (Hathocat, 2013)، وهناك عدة تعريفات للثبات بمعنى الدقة وفقا لطريقة تقدير الثبات والمعادلة المستخدمة ومحددات استخدامها (Worthen et al., 1993).

تركز أدبيات القياس والتقويم على الصدق قبل الثبات في الاختبارات؛ وعند الانتقال من الصدق الى الثبات هناك افتراض بأن مؤشرات الصدق متوفرة، وخاصة صدق محتوى الاختبار القائم على تحليل عضو هيئة التدريس لعناصر المحتوى الدراسي وصياغة الأهداف وصناعة الأسئلة، وحسن اختيار عينة الأسئلة وتمثيلها للمحتوى. وبالمقابل، فهناك الكثير من العوامل ومصادر الأخطاء في مرحلة التخطيط والبناء والتطبيق والتصحيح التي تشكك بصدق هذه الاختبارات وثباتها؛ بما في ذلك الاختلاف بين أعضاء هيئة التدريس في الجامعات، من حيث صفة التساهل والتشدد في إعطاء الدرجات للطلبة في الشعب المختلفة للمقرر الواحد، وربما تتفاعل هذه الخاصية مع جنس الطالب وجنس عضو هيئة التدريس، (Blight, 1988). وينعكس نقص الثبات في الاختبارات الجامعية سلباً على دقة النتائج المبنية عليها، ودقة التوقعات مما يؤكد أهمية تقديم مؤشرات على ثبات الاختبارات وأدوات التقويم الأخرى في خطة المقرر الجامعي. ويصعب تقديم هذه المؤشرات النادرة في أي جامعة، ولا تطلبها الجامعات من أعضاء هيئة التدريس فيها، وكأن الثبات متحقق تلقائيا، وهذا مجرد افتراض، ويعود ذلك الى القدرة والجهد والوقت لتجهيز مصفوفة الدرجات التفصيلية لكل اختبار حسب طريقة حساب الثبات أو معادلة التقدير المستخدمة.

وعلى الرغم مما تعانيه الاختبارات من إعداد المدرسين من مشكلات، إلا أنها متأصلة في العملية التعليمية في الجامعات، وهي الوسيلة المستخدمة لتحديد مدى كفاءة الطالب بالمستويات الأكاديمية خلال البرنامج الأكاديمي للتنبؤ بقدرة الطالب على النجاح والاستمرار في دراسته المستقبلية، وهي من ضمن المعايير المستخدمة في الحكم على جودة المؤسسة التعليميةً. كما ويشكل رصد العلامات وإعداد التقارير حولها غرضاً من أغراض التقويم، حيث تحمل العلامات معانٍ ومدلولات مختلفة لتخدم تلك الأغراض والقرارات التربوية (Gronlund & Robert, 1990). ولذلك يتطلب أن تتمتع العلامات الجامعية في المقررات المختلفة بدرجة مقبولة من الثبات، ويقع على عاتق مؤسسات التعليم مسؤولية تقدير الثبات لعلامات الطلبة في المقررات التي تطرحها في برامجها المختلفة. فالعلامات كانت على الدوام محط الكثير من الانتقادات، والتي من أهمها أن نظم التقدير تخضع لأمور ذاتية وتفتقر إلى الموضوعية والشمول (المسند، ١٩٩٢). وتؤكد أدبيات القياس والتقويم على ضرورة دراسة الثبات في علامات الطلبة لأنه مؤشر كمي على أخطاء القياس، فضعف الثبات يعني ضعف دقة قياس الاختبار للمجال موضع الاهتمام (Wells & Wollaack, 2003).

وكما أن هناك انواع من الصدق في الاختبارات، هناك أنواع من الثبات، أي أنواع من طرق التقدير للثبات وفقا لطبيعة مصادر الأخطاء، والبيانات المتوفرة أو التي يمكن توفيرها، وشكل الاختبار أو أدوات التقويم في المقرر خلال الفصل الدراسي. فلكل طريقة متطلباتها، وهي متطلبات قد تكون صعبة ولا تقع ضمن الامكانات المتاحة، ومنها ما يتطلب حسابات ليست سهلة على أعضاء هيئة التدريس، ولذلك هناك القليل جدا من أعضاء هيئة التدريس قادرين على تقديم مؤشرات ثبات رقمية على دقة تقويمهم لطلبتهم خلال الفصل الدراسي؛ فمن هذه الطرق على مستوى الاختبار الواحد ما يحتاج إلى تطبيق وإعادة تطبيق الاختبار test-retest في ذلك المقرر، وهذا غير ممكن على أرض الواقع؛ ولا يختلف الحال كثيراً إذا استخدمنا طريقة تقدير ثبات التكافؤ لصعوبة تحقيق شروط التكافؤ السوالمة، ١٩٩٥). وكذلك التجانس الداخلي الذي يتطلب مصفوفة درجات تفصيلية على مستوى الفقرة الواحدة لكل طالب، وهذا قد يتوفر عند تطبيق الاختبارات الكترونيا، ويحتاج الى برمجيات وظروف خاصة يندر تحقيقها. وينطبق ذلك أيضا على الاختبار المكون من عدة أجزاء متشاكلة Congeneric (تقيس المجال نفسه، وبينها علاقات ارتباطية قوية)، أو عدة اختبارات متشاكلة خلال الفصل الدراسي، وهي تتطلب إجراءات إحصائية غير مألوفة لأعضاء هيئة التدريس، وصعبة التطبيق. فقد قدم علماء القياس والتقويم الكثير من المعادلات والأساليب المستخدمة في حساب الثبات، وكثرة المساهمين من علماء القياس وما أضافه كل منهم على معادلة غيره، وعُرفت فيما بعد في أدبيات القياس باسم كل منهم مثل Raju, Feldt, Flangan, Angoff-feldt, Kristof, Feldt – Gilmer, Joreskog) أنظر Linn, 1989, P115). ففي عام ١٩٧٧ قام راجو Raju, 1977) بتعميم طريقة كرونباخ ألفا لتقدير ثبات التجانس الداخلي للعلامة الناتجة من قياسات متشاكلة والتي تشترط أن تكون الأوزان النسبية لهذه القياسات معلومة، ومن الجدير بالملاحظة هنا أن معادلة راجو التي تشكل امتدادا لطريقة كرونباخ تتعامل مع خاصية التشاكل، ولا تتعامل مع التباين المشترك في حساب الوزن النسبي للجزء. وتركت الباب مفتوحا لتحديد الأوزان النسبية الداخلة في المعادلة. وإذا نظرنا لمكونات هذه المعادلة، فإنه من السهولة استخدامها مع العلامات الجامعية التي تتوفر لنا من المقررات، باعتبار أن العلامة الجامعية تتكون مجموعة من أجزاء تشمل الاختبارات مثل الاختبار الأول والثاني، والاختبار النهائي، والأعمال الفصلية والمشاركة وغيرها، وهذه جميعها عرضة لأن تختلف في المتوسطات، والتباينات الحقيقية، وتباينات أخطاء القياس، وهو ما تسمح به هذه المعادلة؛ وبالتالي

سهولة الدفاع عن استخدمها في هذه الدراسة، كونها أقرب إلى الواقعية في هذه العلامات.

أما معادلة فلدت – جيلمر Feldt-Gilmer فقد نهجت أسلوبا إحصائيا مختلفا في تحديد الأوزان واشترطت التعامل مع التباينات المشتركة بالإضافة الى التباينات، وهي تتطلب عدة خطوات لحساب هذه الأوزان ليست من السهولة بمكان على أعضاء هيئة التدريس مالم تتوفر برمجيات وتسهيلات تقنية تشجعهم على المبادرة، وهذا غير متوفر في أرض الواقع، ويندر قيام أي عضو هيئة تدريس بالتعرّف على دقة ثبات) عملية التقويم والأدوات التي يستخدمها على مدار الفصل الدراسي. خاصة وأن أدبيات القياس تحدد قيما رقمية ودرجات قطع لمعاملات الثبات المقبولة وغير المقبولة في مستويات متعددة على مبدأ النسبية في الدقة كمعايير للحكم على دقة الاختبارات وأدوات التقويم. ويشير قسم الابتكار التعليمي والتقييم في جامعة تكساس في الولايات المتحدة الأمريكية (DIIA, 2003 إلى عدة درجات قطع في تفسير قيم معاملات الثبات، حيث اعتبرت الثبات المنخفض أقل من٠,٧٠ ، وأن أقل من ٠,٥٠ غير مقبول مطلقا.

وتؤكد الأدبيات على أن تقدير الثبات للعلامات الجامعية يشكل مشكلة سيكومترية معقدة (Sawyer, 1989 نظراً لتعقيد الموقف التعليمي، وكثرة الممارسات التي تطلبها عملية التقويم بالاختبارات، وضمن الإمكانات المتاحة من طرق تقدير الثبات. فقبل ظهور تعميم راجو لمعامل الثبات ألفا، تم استخدام معادلة مشتقة من معادلة سبيرمان براون لتلاءم تقدير ثبات العلامات الجامعية في عينة من المقررات الدراسية قام بها الباحثون ايتاف وايتاف، وهيرد Etauagh, Etaught and Hurd, (1972) توصلوا من خلالها إلى أن معاملات الثبات للمقررات الجامعية متدنية؛ إذ حصلوا على تقديرات تراوحت بين ٠,٣٦ إلى ٠,٤٤،. وبنفس الأسلوب توصلت دراسة شونفلدت وبرش إلى أن ثبات العلامات للمقررات في اثني عشر تخصصاً تراوح بين ٠,٣٩ إلى ٠,٧٦ (Schoenfeldt and Brush, 1997).

واستناداً إلى أبحاث راجو (Raju, 1977) بتعميم طريقة كرونباخ ألفا لتقدير ثبات التجانس الداخلي للعلامة الناتجة من قياسات متشاكلةCongeneric) والتي تشترط أن تكون الأوزان النسبية لهذه القياسات معلومة قام السوالمة (١٩٩٥) بدراسة استخدم فيها معادلة راجو وهدفت إلى تقدير ثبات العلامات الجامعية في عينة من المساقات في جامعة اليرموك، وتكونت بيانات الدراسة من علامات ١٤٥٨ طالباً وطالبة في ٤٨ مساق من مستويات دراسية مختلفة سنة أولى، سنة ثانية، سنة ثالثة، سنة رابعة وفي أربع كليات جامعية هي الآداب، العلوم، الاقتصاد، التربية والفنون. وقد أشارت نتائج الدراسة إلى تمتع العلامات في ٥٦٪ من المساقات الجامعية بدرجة مقبولة من الثبات، حيث يزيد معامل الثبات لها عن٠,٧٠. وتميل العلامات في كليات الآداب والاقتصاد لأن تكون أعلى ثباتاً في المتوسط من العلامات في كليات العلوم والتربية والفنون. أما فيما يتعلق بالمستوى الدراسي فلم تشر النتائج إلى وجود علاقة بينه وبين مستوى الثبات على الرغم من أن متوسط معامل الثبات للعلامات الجامعية في مستوى السنة الأولى أعلى منه في السنوات الجامعية التالية.

وبنفس أسلوب دراسته السابقة قام السوالمة (٢٠٠١) بدراسة مشابهة أخرى لتقدير معاملات الثبات في ١٠٨٧ شعبة دراسية لمساقات كلية العلوم الإنسانية والاجتماعية في جامعة الإمارات العربية المتحدة، ومحاولاً تقصي تأثير تحليل نتائج الطلبة على تقدير معاملات الثبات حيث تم رصد علامات ٥٣٠ شعبة دراسية في الفصل الأول قبل إجراء تحليل لنتائج الطلبة، و٥٥٧ شعبة دراسية في الفصل الثاني بعد إجراء تحليل لنتائج الطلبة، واستخدمت معادلة راجو لتقدير ثبات الاتساق الداخلي للعلامات الجامعية، وتبين أن العلامات الجامعية تتمتع بمستوى معقول من الثبات في كلا الفصلين، حيث تبين أن ٥٨٪ من المساقات تتمتع بمعامل ثبات ٠,٧ وأكثر، وتبين

وجود فروق توزيعات الثبات بين الفصلين لصالح الفصل الثاني، حيث ارتفعت نسبة المساقات ذات الثبات المرتفع من ٥٣٪ في الفصل الأول إلى ٦٢٪ في الفصل الثاني، وانخفضت نسبة المساقات ذات الثبات المنخفض من ١١٪ في الفصل الدراسي الأول إلى ٨٪ في الفصل الدراسي الثاني.

وعلى غرار دراسة السوالمة قام الشايب٢٠٠٧) بدراسة هدفت إلى تقدير ثبات العلامات الجامعية لعينة مكونة من ٦٤ مادة في جامعة آل البيت، ومدى اختلاف تقديرات الثبات باختلاف مستوى المادة ، واختلاف الكلية.. وقد كشفت نتائج استخدام أسلوب راجو (Raju, 1977) لتقدير الثبات عن تدني معاملات ثبات العلامات بشكل عام، حيث بلغ المتوسط العام لتقديرات الثبات ٠,٦٦. كما أشارت النتائج إلى أن ٣١.٢٥٪ فقط من المواد تتمتع بمعاملات ثبات مقبولة. وأشارت نتائج الإحصائي مربع كاي للاستقلالية إلى عدم وجود علاقة بين ثبات علامة المادة ومستوى المادة أو الكلية التي تنتمي إليها.

ومن خلال دراسة تحليلية سيكو مترية لعلامات الطلبة في عينة من المقررات في جامعة القدس المفتوحة قام بها أبو جراد (٢٠٠٩) تكونت بيانات الدراسة من علامات ٤١١٥ طالبا وطالبة موزعين على ٥٢ شعبة من مختلف المستويات الدراسية في أربعة برامج أكاديمية، وأشارت النتائج إلى تمتع العلامات بدرجة مقبولة من الثبات في ٨٨٪ من المساقات حيث زاد معامل الثبات لها عن ٠,٧٠. كما أشارت نتائج الإحصائي كاي تربيع إلى استقلالية معاملات الثبات عن كل من: البرنامج الدراسي الذي ينتمي إليه المقرر والمستوى الدراسي للمقرر.

وفي دراسة قام بها الأحمد (٢٠١٠) هدفت إلى تقدير معامل الثبات للعلامات الجامعية في جميع مساقات البكالوريوس في جامعة جدارا في الأردن، لعينة مكونة من ١٢٤ شعبة من مختلف الكليات، وأشارت النتائج أن ٣٣,٠٧٪ من المساقات كان ذات ثبات منخفض، ٣١,٤٥٪ من المساقات ذات ثبات متوسط، ٣٥,٤٨٪ من المساقات ثباتها مرتفع، واختلفت النتائج باختلاف الكلية.

وحديثاً قام كل من العمري والعكور (٢٠١٤)، بدراسة هدفت إلى الكشف عن أثر حجم الشعبة الدراسية على تقديرات الثبات باستخدام معادلة راجو لثلاثة أجزاء فأكثر معروفة الأوزان. على عينة من ٦٣ شعبة دراسية عشوائيا بجامعة جدار في الاردن بواقع ٢١ شعبة لكل من الشعب الصغيرة والمتوسطة والكبيرة. وأشارت نتائج الدراسة بشكل عام إلى انخفاض مستوى الثبات للعلامات الجامعية حيث وصل المتوسط العام إلى٠,٥٥)، وأن ٢٥٪ من الشعب كانت منخفضة الثبات و٦٥٪ منها غير مقبولة الثبات. وأن متوسط الثبات في الشعب الصغيرة ٠,٦٨ ، فيما انخفض إلى ٠,٤١ في الشعب الكبيرة.

وعلى مستوى تقدير الثبات على العلامات المدرسية الناتجة من اختبارات من إعداد المعلمين، أجريت دراسة في مدينة إربد في الأردن قامت بها عليان (٢٠٠٣)، وهدفت إلى تقدير الثبات في عينة من ٢٠٧ شعبة دراسية في مدارس مدينة إربد، واستخدمت معادلة راجو لتقدير الثبات في المواد الدراسية، وأظهرت نتائج الدراسة ارتفاع معاملات الثبات للمواد الدراسية؛ إذ تبين أن ٩٢,٣٪ يزيد معامل ثباتها عن ٠,٨

ويتضح من الدراسات السابقة أن تقدير الثبات لعلامات الطلبة يشكل كل منها دراسة حالة. لأن مصادر الأخطاء متعددة كما ونوعا، مما يعني ضرورة اهتمام الجامعات بتمكين أعضاء هيئة التدريس من تقديم مؤشرات على دقة اختبارات وأساليب التقويم الواردة في خطط المقررات، حتى لا تكون العلامات الجامعية مجرد أرقام؛ ولذلك يتوقع زيادة الاهتمام بتوفير الفرص وتقديم التسهيلات التي تشجع أو تدفع أعضاء هيئة التدريس لفحص ثبات أدوات التقويم. حتى لو كلفهم ذلك جهدا زائدا ضمن حدود المعقول بالتزامن مع تدريبهم على إجراءات الحصول على مؤشرات رقمية للثبات.

لتمكينهم من مراجعة أدوارهم التدريسية ومنها الدور التقويمي حتى لا يبقى تصورهم لدقة تقويمهم قائم على افتراضات خاطئة في غياب المؤشرات الرقمية على مستوى الثبات.

في ضوء هذا التصور جاءت هذه الدراسة لاستكشاف واقع دقة علامات الطلبة في عينة من المقررات وتحديدا عينة المقررات التي تطرحها السنة التحضيرية في جامعة تبوك، ويتوقع أن تكون من الدراسات السباقة في إطار تقدير ثبات العلامات الجامعية في الجامعات السعودية باستخدام تعميم راجو لتقدير ثبات الاتساق الداخلي على المكونات أو الأجزاء، ومعادلة راجو المعدّلة من حيث طريقة التعامل مع الأوزان النسبية اعتمادا على مؤشر الانحرافات المعيارية ومقارنة نتائجها بالثبات المقدّر باستخدام أوزان محددة مسبقا في التعليمات، ومقارنتها ايضا بالتقديرات من معادلة فيلدت - وجيلمر التي تفترض وجود تباينات مشتركة بين الأجزاء، بغرض التعرف على درجة تحمل معادلة راجو المعدّلة لانتهاك هذا الافتراض، الذي قد يدفع أعضاء هيئة التدريس لفحص دقة اختباراتهم بسهولة من بيانات متوفرة وبدرجة مقبولة من الدقة في التقدير للثبات. مما قد يشكل ذلك إضافة نوعية في إطار المعايير الناعمة لجودة التعليم الجامعي.

مشكلة البحث

الثبات في الاختبارات كما أشرنا في المقدمة تعني دقة قياس تلك الاختبارات لما تدعي قياسه، على افتراض أن باني الاختبار تجاوز الحديث عن خاصية الصدق، وأن لديه ما يكفي من مؤشرات على صدقه، مع أن تحقيق هذه الخاصية في الاختبارات من اعداد عضو هيئة التدريس بشكل خاص يبقى مجرد افتراض. وغالبا ما يلتزم عضو هيئة التدريس بالأوزان النسبية التي تحددها تعليمات الجامعة عند إعدادهم لخطط المقررات التي تتضمن التقويم لنواتج التعلم بالاختبارات وبغير الاختبارات، ومنها مقررات السنة التحضيرية بجامعة تبوك. وبمتابعة خصائص الدرجات على الاختبارات خلال الفصول المتلاحقة في شعب الطلاب والطالبات المختلفة لكل مقرر (أي وفق متغيري الشعبة والشطر)، لوحظ من خلال التحليل الدوري لتلك الاختبارات ونتائجها أن التعامل مع النسب التي تحددها التعليمات تفرز معاملات ثبات ملوثة احصائيا وغير دقيقة، فقد يكشف الاختبار الأول عن فروق حقيقية بين الطلبة بينما لا يكشف النهائي عن أي فروق، مع أن له وزنا نسبيا يساوي ضعف الاختبار الأول؛ وهذا يعني أن الأوزان النسبية التي تحددها التعليمات ليس بالضرورة أن تكون متوافقة مع الأوزان التي تعتمدها المعادلات المستخدمة في تقدير الثبات، ومنها معادلة راجو Raju القائمة على:

- وجود نماذج متشاكلة من الاختبار Congeneric Forms على افتراض أنها أجزاء Parts.
- عدد الفقرات أو الأسئلة محددة في كل جزء Part ويمكن أن تكون متساوية كحالة خاصة.
- يعطى كل قسم وزنا نسبيا يساوى مقلوب عدد الفقرات Items.
- لا تدخل فيها حسابات قائمة على التباينات المشتركة covariance للأجزاء.
- تعتمد كليا في الأوزان على عدد الفقرات فقط بصرف النظر عن خصائص الفقرة الواحدة.

ومنها أيضاً معادلة فلدت – جيلمر Feldt-Gilmer القائمة على:

- أوزان نسبية للأجزاء تأخذ بالاعتبار التباينات Variance والتباينات المشتركة Covariance.
- ان الدرجة الكلية على الاختبار تساوى المجموع للدرجات على الأجزاء.
- تتطلب فقط الانحراف المعياري للدرجات الكلية على الاختبار، والانحرافات المعيارية

للدرجات على الأجزاء ، ولا تتطلب ذلك على مستوى الفقرة الواحدة وهذا ضمن حدود إمكانات أعضاء هيئة التدريس في الواقع العملي).

لكل هذه الأسباب مجتمعة سواء من حيث شروط أو افتراضات استخدام تلك المعادلات، والامكانات المتوفرة لاستخدامها في البيئة الجامعية، والرغبة في تقديم مؤشرات سيكومترية على خصائص الاختبارات الجامعية في إطار الجودة، بالإضافة الى ما تشير اليه الأدبيات من أهمية الانحراف المعياري كمؤشر إحصائي على مستوى الفقرة الواحدة في الاختبار، والذي يحدد معامل تمييز الفقرة، وكذلك أهميته في تقدير الثبات، لأن أي فقرة أو جزء من الاختبار، أو الاختبار بأكمله الذي لا يميز بين الطلبة، فمن الطبيعي أن يكون ثباته كمؤشر دقة ضعيفا. في ضوء هذا التصور القائم على التحليل المنطقي والاحصائي لأهمية الثبات في الاختبارات الجامعية من جهة، والمحددات التي تواجه أعضاء هيئة التدريس في فحص اختباراتهم ضمن الامكانات المتاحة، والتعامل مع الحد الأدنى من الافتراضات في القديرات من جهة أخرى؛ جاءت المشكلة في هذا البحث، فهو بحث يجمع بين الجانب النظري والميداني، حيث أن معايير الجودة لأدوار أعضاء هيئة التدريس بحاجة الى فحص خصائص اختباراتهم بطريقة علمية وميسرة (Feasible) – وهذا هو الهدف الرئيسي لهذا البحث، والمتمثّل بتقديم معادلة جديدة قائمة على الانحرافات المعيارية كأوزان نسبية لأجزاء الاختبار الواحد من الاختبارات الجامعية على مدار الفصل على افتراض أن كل اختبار هو جزء من بطارية اختبارات وأدوات التقويم خلال الفصل الدراسي)، وهو هدف لم تتطرق اليه أدبيات القياس والباحثين في المعادلات المستخدمة لتقدير الثبات حسب المعادلتين المشار اليهما أو غيرهما من المعادلات، وهذا لا يمنع من أن تكون هذه المعادلة كغيرها مع المعادلات قائمة على بعض المحددات العملية أو بعض الافتراضات، مثل تحملها لانتهاك افتراض التباينات المشتركة بين الأجزاء الذي تفترضه بعض المعادلات في حساب الأوزان النسبية كما هو الحال في معادلة فيلدت - جيلمر. وعلى أي حال، فهي مساهمة تتطلب المزيد من الدراسات، لفحص إمكانية تجاوز الافتراضات الاحصائية ودرجة تحملها Robustness لهذا التجاوز. خاصة وأن بعض المعادلات تقوم في الأوزان النسبية التي تستخدمها على عدد الفقرات، وهذا أيضا قائم على افتراضات يصعب تحقيقها مثل تساوي صعوبة هذه الفقرات وقدرتها التمييزية، كما هو الحال في معادلة راجو. ولذلك فإن معادلة راجو المعدّلة والتي يمكن أن يشار اليها بمعادلة عودة -جودة Audeh-Judeh reliability Formula) يمكن أن تشكل إضافة نوعية لأدبيات القياس والتقويم في هذا الاطار.

أسئلة الدراسة

في ضوء التصور السابق لمشكلة الدراسة، جاءت الأسئلة التالية لوصف واقع الاختبارات لمقررات السنة التحضيرية في جامعة تبوك وفق متغيري الشعبة والشطر من جهة، باستخدام معادلة راجو باعتماد الأوزان النسبية للاختبارات في تعليمات الجامعة مرة، ثم القائمة على الانحرافات المعيارية مرة أخرى، مقابل معاملات الثبات المقدّرة بمعادلة فيلدت – جيلمر القائمة على التباينات والتباينات المشتركة:

السؤال الأول: ما دقة ثبات علامات الطلبة في مقررات السنة التحضيرية بجامعة تبوك مقدرّة بمعادلة راجو المعدّلة التالية؟:

$$\text{Raju-Audeh-Judeh } RAJ_{w\%} = \left\{\frac{1}{1-\sum w_{\%}^2}\right\}\left\{1-\frac{\sum s_t^2}{s_T^2}\right\}$$

وذلك باعتماد أوزان قائمة على النسب المخصصة لأدوات التقويم في تعليمات الجامعة لائحة الدراسة والاختبارات.

السؤال الثاني: هل يختلف توزيع تقديرات الثبات المقدّرة بمعادلة $RAJ_{w\%}$ باختلاف متغيري المقرر، والشطر (طلاب، طالبات)؟

السؤال الثالث: و ما دقة ثبات) علامات الطلبة في مقررات السنة التحضيرية بجامعة تبوك مقدرّة بمعادلة راجو المعدّلة التالية؟:

Raju-Audeh-Judeh $RAJ_{ws} = \{\frac{1}{1-\sum w_s^2}\}\{1-\frac{\sum s_t^2}{s_T^2}\}$

باعتماد الانحرافات المعيارية للدرجات على الاختبار الواحد كأوزان نسبية.

السؤال الثالث: ما دقة ثبات) علامات الطلبة في مقررات السنة التحضيرية بجامعة تبوك مقدرّة بمعادلة فيلدت – جيلمر القائمة في حساب الأوزان النسبية على مصفوفة التباينات والتباينات المشتركة(variance –covariance matrix

$$\text{Feldt–Gilmer} = ((\Sigma D_i)^2/(\Sigma D_i)^2 - \Sigma D_i^2)((\sigma_x^2 - \Sigma\sigma_i^2)/\sigma_x^2);$$

$$D_i = (\Sigma\sigma_{ij} - \sigma_{it} - \sigma_i^2)/(\Sigma\sigma_{ij} - \sigma_{il} - \sigma_i^2)(\text{sum is over j}),$$

(ملاحظة: للتعرف على طريقة تقدير الثبات بعد انشاء مصفوفة التباينات والتباينات المشتركة باستخدام هذه المعادلة يمكن الرجوع الى (Linn,1989, p. 110-116)

السؤال الرابع: هل يوجد فروق ذات دلالة احصائية أو عملية بين متوسطات التقديرات لمعاملات الثبات المحسوبة بالمعادلات الثلاث؟ ملاحظة: معادلة فلدت – جيلمر هي المرجعية في المقارنة لأنها لا تتجاوز التباينات المشتركة، علما بأن معادلة راجو الأصلية لا تتعامل مع التباينات المشتركة بين أجزاء الاختبار، مع أنها موجودة؛ ولذلك يبقى فحص مدى تأثر دقة التقديرات التي تأخذ بالاعتبار التباينات المشتركة في التقدير مقابل المسوغات العملية المشجعة لعضو هيئة التدريس وامكانية التطبيق مع الحد الأدنى من الدقة هو الأساس في الاختيار لطريقة التقدير).

مجتمع الدراسة وعينتها

استخدمت هذه الدراسة أسلوب المسح الشامل لجميع مقررات التحضيرية لتقدير الثبات عند استخدام معادلة راجو المعدّلة، وأسلوب العينة عند استخدام معادلة فيلدت وجيلمر (نظرا لكثرة الحسابات والعمليات الاحصائية التي يتطلبها حساب التقدير للثبات على مستوى كل شعبة/مقرر، وعدم وجود برمجيات لهذا الغرض)، حيث بلغ عدد المقررات خمسة عشر مقرراً بواقع ٦٦٣ شعبة دراسية في شطري الطلاب والطالبات، كما هو مبين في جدول ١.

جدول ١

أعداد الشعب موزعة حسب الشطر والمقرر

اسم المقرر	عدد الشعب للطلاب	عدد الشعب للطالبات	العدد الكلي
الأحياء	١٦	١٧	٣٣
الثقافة الإسلامية	١٣	١٦	٢٩
الحاسب الآلي	٤٣	٤٧	٩٠
الرياضيات ١	٩	٥	١٤
الرياضيات ٢	٢١	٢٥	٤٦
الفيزياء	١٨	١٨	٣٦
الكيمياء	١٦	١٧	٣٣
انجليزي ١	٢٠	١٣	٣٣
انجليزي ٢	٥٥	٧١	١٢٦
مبادئ الإدارة	١٣	١٧	٣٠
مبادئ الرياضيات	١٣	١٤	٢٧
مقدمة في الإحصاء	١٢	١٦	٢٨
مهارات الاتصال	٢٧	٢٨	٥٥
مهارات التفكير والبحث	٢٦	٢٩	٥٥
مهارات لغوية	١٢	١٦	٢٨
المجموع	٣١٤	٣٤٩	٦٦٣

حدود الدراسة ومحدداتها

- استخدام الأوزان حسب النسب من مئة التي تحددها الجامعة لاختبار الدوري الأول والدوري الثاني والنهائي) في معادلة راجو في تقدير معاملات الثبات للعلامات الجامعية لجميع الشعب في المقررات.

- استخدام الأوزان النسبية القائمة على الانحرافات المعيارية للدرجات الواقعية على الاختبارات كصورة معدّلة لمعادلة راجو في تقدير معاملات الثبات للعلامات الجامعية لجميع الشعب في المقررات.
- تقدير معاملات الثبات باختيار عينة عشوائية٥٪) من جميع الشعب عند استخدام معادلة فيلدت – جيلمر
- اعتماد علامات اختبار الدوري الأول والدوري الثاني والنهائي في تقدير الثبات، واستبعاد علامة المشاركة من ذلك، نظراً لعدم وجود معايير واضحة لدى المدرسين في تقديرها للطالب.

المعالجات الإحصائية

تتلخص الاجراءات الاحصائية التي تجيب عن أسئلة الدراسة بما يلي:

- تقدير الثبات باستخدام معادلة راجو (Raju) في جميع الشعب لكل مقرر وفق الأوزان النسبية المحددة في لائحة الدراسة والاختبارات في الجامعة، وليس وفق نسب أطوال الاختبارات مقدرّة بعدد الأسئلة في كل اختبار أو جزء الى العدد الكلي لأنها غير متوفرة وغير منطقية. ثم استخرجت خصائص توزيع متوسطات الثبات وفق متغيري الشعبة للمقرر، والشطر.
- أعيد تطبيق معادلة راجو على جميع الشعب في المقررات باستخدام الانحرافات المعيارية للاختبارات المكونة للعلامة الكلية كأوزان نسبية في المعادلة.
- تقدير الثبات مرة ثالثة باستخدام معادة فلدت - جيلمر، التي تعتمد التباينات والتباينات المشتركة في حساب الأوزان النسبية على عينة من المقررات٣٠مقرر) بواقع شعبة عشوائيا من كل مقرر.
- وبعد استخراج معاملات الثبات بالطرق الثلاث، تم مقارنة متوسطات تقديرات الثبات المقدّرة بطريقة الأوزان القائمة على التعليمات بطريقة أوزان القائمة على الانحرافات المعيارية على جميع الشعب وعددها ٦٦٣ شعبة)، وذلك من خلال استخدام اختبار T للعينات المرتبطة Paired Sample T-test، كما تم مقارنة متوسطات الثبات بالطرق الثلاث لتحديد الأوزان التعليمات، والانحرافات المعيارية، ومصفوفة التباينات والتباينات المشتركة) باستخدام تحليل التباين للقياسات المتكررة.

النتائج

بعد حساب الانحرافات المعيارية لعلامات الطلبة على الاختبارات الثلاثة في جميع المقررات، تم تطبيق معادلة راجو (Raju) على درجات الطلبة بعد حساب التباينات الكلية والفرعية على الاختبارات الثلاثة لكل مقرر، ومربعات أوزان مكونات العلامة الكلية. وتبين أن متوسط الثبات في جميع المقررات لكل الشعب هو ٠,٦٨ بانحراف معياري قدره ٠,١٤، فيما تراوحت قيم معاملات الثبات في الشعب في جميع المقررات بين ٠,١٦ إلى ٠,٩٦، ثم قدّر الثبات في كل مقرر، وحسب نوع الشطر، وسجلت نتائج تقدير متوسطات الثبات في الجدول رقم ٢ للإجابة عن السؤال الأول.

وقد تم تصنيف معاملات الثبات في جميع المقررات في فئات الثبات وفق معيار (DIIA, 2003 بعد إعادة التصنيف في ثلاث فئات. وتبين أن ما نسبته ٥٧٪ من المقررات كانت درجة ثباتها متوسطة، مقابل ١٨٪ مرتفعة ثبات أعلى من ٨٠.). ولفحص العلاقة بين مستوى الثبات ونوع المقرر والشطر، تم بناء جدول التوافق (Cross Tabulation بين متغير مستوى الثبات، ونوع المقرر، وسجلت النتائج في جدول ٣.

جدول ٢

متوسطات الثبات لشعب المقرر موزعة حسب الشطر والعينة ككل٦٦٣شعبة) محسوبة بمعادلة راجو وفق أوزان نسبية محددة في التعليمات

المقرر	متوسط الثبات (الطلاب)	الانحراف المعياري	متوسط الثبات الطالبات)	الانحراف المعياري	متوسط ثبات جميع المقررات	الانحراف المعياري
الأحياء	٠,٦١	٠,١٣	٠,٧٨	٠,١٣	٠,٧٠	٠,١٥
الثقافة الإسلامية	٠,٦٧	٠ ,١١	٠,٧٢	٠,١١	٠,٧٠	٠,١١
الحاسب الآلي	٠,٦٥	٠,١٧	٠,٦٧	٠,١٥	٠,٦٦	٠,١٥
الرياضيات ١	٠,٥٣	٠,١٢	٠,٦١	٠,٢٣	٠,٥٦	٠,١٦
الرياضيات ٢	٠,٦٦	٠,١٤	٠,٧٥	٠,٠٩	٠,٧١	٠,١٢
الفيزياء	٠,٧٧	٠,١٠	٠,٨٠	٠,٠٩	٠,٧٩	٠,٠٩
الكيمياء	٠,٦٢	٠,١٠	٠,٧٧	٠,١٢	٠,٧٠	٠,١٣
انجليزي ١	٠,٦٢	٠,١١	٠,٦٧	٠,١٦	٠,٦٤	٠,١٣
انجليزي ٢	٠,٦٦	٠,١٦	٠,٦٩	٠,١٠	٠,٦٨	٠,١٣
مبادئ الإدارة	٠,٦٢	٠,١٠	٠,٧٧	٠,٠٩	٠,٧٠	٠,١٢
مبادئ الرياضيات	٠,٥٧	٠,١٣	٠,٨٠	٠,٠٧	٠,٦٩	٠,١٥
مقدمة في الإحصاء	٠,٦٥	٠,١٥	٠,٧٨	٠,١١	٠,٧٢	٠,١٤
مهارات الاتصال	٠,٦٢	٠,١٧	٠,٦٠	٠,١٣	٠,٦١	٠,١٤
مهارات التفكير والبحث	٠,٦٠	٠,١٤	٠,٦٧	٠,١٠	٠,٦٣	٠,١٢
مهارات لغوية	٠,٧٣	٠,٠٦	٠,٦٧	٠,١٠	٠,٦٩	٠,٠٩
المقررات ككل	٠,٦٤	٠,١٤	٠,٧١	٠,١٣	٠,٦٨	٠,١٤

وأشارت نتائج اختبار كاي تربيع للاستقلال إلى وجود اختلاف جوهري في معاملات الثبات تعزى الى نوع المقرر؛ حيث بلغت قيمة كاي تربيع ٩٨,٦، وهذه القيمة ذات دلالة إحصائية عند مستوى دلالة أقل من ٠,٠٠١ وهذه النتيجة تعني اختلاف عدد المقررات ونسبتها في كل مستوى من مستويات الثبات باختلاف نوع المقرر، ولفحص العلاقة بين مستوى الثبات ومتغير الشطر، تم بناء جدول التوافق (Cross-Tabulation) بين متغير مستوى الثبات، ونوع الشطر، وسجلت النتائج في جدول ٤.

وأشارت نتائج اختبار كاي تربيع للاستقلال إلى وجود اختلاف جوهري أيضاً بين نوع الشطر ومستوى الثبات؛ إذ بلغت قيمة كاي تربيع ٣٥,٣، وهذه القيمة ذات دلالة إحصائية عند مستوى دلالة إحصائية أقل من ٠,٠٠١ وهذا يعني اختلاف نسب المقررات حسب نوع الشطر.

وفي إطار الاجابة عن السؤال الثالث، تم إعادة تطبيق معادلة راجو(Raju على درجات الطلبة، ولكن باستخدام الانحرافات المعيارية كأوزان نسبية، وقدّرت متوسطات الثبات لكل شعب المقررات وفي المقررات ككل، وسجلت النتائج في جدول ٥.

جدول ٣

أعداد المقررات موزعة حسب نوع المقرر ومستوى الثبات

المقرر	مستوى الثبات		
	مرتفع	متوسط	منخفض
مهارات لغوية	٤	٢٢	٢
رياضيات ١	٩	٣	٢
مهارات التفكير والبحث	١٦	٣٥	٤
الأحياء	٧	١٧	٩
الكيمياء	٧	١٧	٩
الرياضيات ٢	٦	٢٨	١٢
الفيزياء	٣	١١	٢٢
الحاسب الآلي	٢٢	٥٥	١٣
مهارات اتصال	٢١	٣٢	٢
انجليزي ١	١٠	٢١	٢
انجليزي ٢	٢٨	٧٩	١٩
مقدمة في الإحصاء	٥	١٣	١٠
الثقافة الإسلامية	٧	١٨	٤
مبادئ الإدارة	٨	١٢	٧
مبادئ الرياضيات	٦	١٧	٧
المجموع		٣٨٠	١٢٤

جدول ٤

أعداد المقررات موزعة حسب مستوى الثبات والشطر

مستوى الثبات		الشطر		
		طلاب	طالبات	المجموع
مستوى الثبات	متدني	٩٥	٦٤	١٥٩
	متوسط	١٨٨	١٩٢	٣٨٠
	مرتفع	٣١	٩٣	١٢٤

وتشير هذه النتائج إلى ارتفاع متوسطات الثبات المقدّرة بهذه الطريقة، وقد تم حساب معامل الارتباط بين متوسطات الثبات المقدّرة باستخدام الأوزان المحددة في التعليمات، وتلك المحسوبة باستخدام الانحرافات المعيارية، وتبين وجود علاقة قوية بينهما إذ بلغ ٠,٩٤ وأشارت نتائج مقارنة متوسطات تقديرات الثبات المقدّرة بالطريقتين باستخدام اختبار T للعينات المرتبطة Paired Sample T-test إلى أن غالبيتها دالة إحصائياً والتي تظهر في جدول ٦.

جدول ٥

متوسطات الثبات في المقررات باستخدام الأوزان الإحصائية للانحرافات المعيارية للاختبارات في المقررات

المقرر	متوسط الثبات	الانحراف المعياري
الأحياء	٠,٧٢	٠,١٣
الثقافة الإسلامية	٠,٧١	٠,١٠
الحاسب الآلي	٠,٦٩	٠,١٥
1الرياضيات	٠,٦٠	٠,١٥
2الرياضيات	٠,٧٢	٠,١١
الفيزياء	٠,٧٩	٠.٠٩
الكيمياء	٠,٧٢	٠,١١
1 انجليزي	٠,٧٠	٠,١٣
2 انجليزي	٠,٧٠	٠,١٢
مبادئ الإدارة	٠,٧٢	٠,٠٩
مبادئ الرياضيات	٠,٧٢	٠,١٤
مقدمة في الإحصاء	٠,٧٤	٠,١٣
مهارات الاتصال	٠,٦٢	٠,١٢
مهارات البحث والتفكير	٠,٦٦	٠,٠٩
مهارات لغوية	٠,٧٠	٠,١٠
المقررات ككل	٠,٧٠	٠,١٣

جدول ٦

نتائج مقارنة متوسطات تقديرات الثبات المقدّرة بطريقة الأوزان التي تحددها التعليمات وتلك المقدّرة بالانحرافات المعيارية، باستخدام اختبار ت للعينات المرتبطة

المقرر	ت	د.ح.	الدلالة الإحصائية
الأحياء	٢.٨٥	٣٢	٠,٠٠٨
الثقافة الاسلامية	٣.٣٣	٢٨	٠,٠٠٢
الحاسب الآلي	٧,٤٩	٨٩	٠,٠٠٠
الرياضيات ١	٥,٠٢	١٣	٠,٠٠٠
الرياضيات ٢	٢,٩٥	٤٥	٠,٠٠٥
الفيزياء	١,١٧	٣٥	٠,٢٥١
الكيمياء	٤,١٤	٣٢	٠,٠٠٠
انجليزي ١	٤,١٧	٣٢	٠,٠٠٠
انجليزي ٢	٤,٠٥	١٢٥	٠,٠٠٠
مبادئ الإدارة	١,٥٦	٢٩	٠,١٣٠
مبادئ الرياضيات	٣,٥١	٢٦	٠,٠٠٢
مقدمة في الإحصاء	٢,٦٩	٢٧	٠,٠١٢
مهارات الاتصال	٢,٥٥	٥٤	٠,٠١٤
مهارات البحث والتفكير	٥,٤٩	٥٤	٠,٠٠٠
مهارات لغوية	٠,٨٦	٢٧	٠,٣٩٩
المقررات ككل	١٢,٠٠	٦٦٢	٠,٠٠٠

ويلاحظ من جدول ٦ أن فروق المتوسطات لمعاملات الثبات المقدّرة بالطريقتين دالة إحصائياً في غالبية المقررات. وللإجابة عن السؤال الرابع القائم على فحص تأثير التباينات والتباينات المشتركة للأجزاء الفرعية المكونة للعلامة الكلية على تقديرات متوسطات الثبات ومقارنتها بالطريقتين السابقتين، تم اختيار عينة عشوائية من المقررات عددها ٣٠ شعبة دراسية، وطبقت عليها معادلة فلدت – جيلمر Feldt-Gilmer، وسُجّلت النتائج في جدول ٧ الذي يحتوي على متوسطات الثبات المقدّرة بالطرق الثلاث راجو القائمة على الأوزان من التعليمات R1، راجو القائمة على الانحرافات المعيارية R2، وفيلدت – جيلمر R3.

وبفحص الفروق بين المتوسطات باستخدام تحليل التباين للقياسات المتكررة، تبين أنه توجد فروق ذات دلالة إحصائية بين التقديرات عند مقارنة الطريقة الأولى (R1) بكل من الطريقتين الثانية (R2) والثالثة (R3)، على مستوى دلالة 001.، وبالمقابل لا يوجد فرق دال إحصائيا بين تقديرات معاملات الثبات بالطريقة الثانية (R2) وتلك المقدّرة بالطريقة الثالثة (R3).

كما تم حساب معاملات الارتباط لتقديرات الثبات بالطرق الثلاث وكانت على التوالي 93. =R1R2، 92.=R1R3، 98.=R2R3)، حيث يلاحظ التوافق العالي جدا بين التقديرات المحسوبة بالطريقتين القائمتين على الانحرافات المعيارية كأوزان نسبية مقترحة من الباحثين في هذه الدراسة لتستخدم بمعادلة راجو بدون معرفة أطوال الاختباراتUnknown Lengths)، مقابل طريقة فيلدت - جيلمر التي تتطلب عمليات إحصائية قائمة على التباينات والتباينات المشتركة.

جدول ٧

متوسطات الثبات المقدّرة بالطرق الثلاث راجو القائمة المعدلة على الأوزان من التعليمات (R1) ، وراجو المعدلة القائمة على الانحرافات المعيارية (R2)، وفيلدت –جيلمر (R3) على عينة مكونة من ٣٠ شعبة دراسية

رقم الشعبة	R1	R2	R3	رقم الشعبة	R1	R2	R3	رقم الشعبة	R1	R2	R3
١	٠,٧٣	٠,٧٣	٠,٧٥	١١	٠,٧٨	٠,٧٦	٠,٧٦	٢١	٠,٦٩	٠,٧٦	٠,٧٤
٢	٠,٧٨	٠,٧٩	٠,٧٧	١٢	٠,٧٨	٠,٧٦	٠,٧٦	٢٢	٠,٧٨	٠,٧٩	٠,٧٧
٣	٠,٧٧	٠,٧٧	٠,٧٦	١٣	٠,٤٩	٠,٥٧	٠,٦٠	٢٣	٠,٥٢	٠,٥٨	٠,٦٠
٤	٠,٧٨	٠,٧٣	٠,٧٣	١٤	٠,٦٤	٠,٦٩	٠,٦٩	٢٤	٠,٨٤	٠,٨٧	٠,٨٩
٥	٠,٧٩	٠,٧٩	٠,٨٠	١٥	٠,٣٦	٠,٤٤	٠,٤٣	٢٥	٠,٦٧	٠,٦٨	٠,٧٠
٦	٠,٣٤	٠,٣٧	٠,٥٥	١٦	٠,٣٨	٠,٤٨	٠,٥٢	٢٦	٠,٥٠	٠,٥١	٠,٦١
٧	٠,٦٥	٠,٦٨	٠,٦٩	١٧	٠,٧٣	٠,٨٣	٠,٨٦	٢٧	٠,٦٠	٠,٧١	٠,٧٢
٨	٠,٧٧	٠,٧٧	٠,٧٦	١٨	٠,٢٦	٠,٣٠	٠,٢٦	٢٨	٠,٦٣	٠,٦٨	٠,٦٨
٩	٠,٦٩	٠,٦٩	٠,٦٨	١٩	٠,٧٥	٠,٧٨	٠,٧٩	٢٩	٠,٨٤	٠,٨٣	٠,٨٢
١٠	٠,٥٩	٠,٦١	٠,٦٠	٢٠	٠,٧٣	٠,٧٢	٠,٧٣	٣٠	٠,٥١	٠.٥٦	٠,٦٠
المتوسط الحسابي									٠,٦٥	٠,٦٧	٠,٦٩
الانحراف المعياري									٠,١٦	٠,١٤	٠,١٣
اقل معامل ثبات									٠,٢٦	٠,٣٠	٠,٢٦
أعلى معامل ثبات									٠,٨٤	٠,٨٧	٠,٨٩

R1 = معاملات الثبات مقدّرة بمعادلة راجو القائمة على الأوزان النسبية للاختبارات المحددة في التعليمات

R2 = المعاملات الثبات مقدّرة بمعادلة راجو القائمة على الانحرافات المعيارية

R3 = المعاملات الثبات مقدّرة بمعادلة فلدت – جيلمر القائمة على مصفوفة التباينات والتباينات المشتركة

مناقشة النتائج

أظهرت نتائج تقدير الثبات باستخدام معادلة $RAJ_{w\%}$ على أن متوسط الثبات في عينة الدراسة من المقررات هو ٠,٦٨ وهي درجة متوسطة بشكل عام، إلا أنها بمعيار نسبة الدقة الى عدم الدقة signal : noise Ratio كنسبة ١:٢ تقريبا يمكن وصفها بأنها ضعيفة نسبيا وخاصة عندما يكون الغرض من الدرجات اتخاذ قرارات مصيرية بحق الطلبة كما هو الحال بالنسبة لطلبة السنة التحضيرية حيث تستخدم الدرجات لأغراض مواصلة الدراسة الجامعية، والتنبؤ في النجاح المستقبلي للتخصص الذي يدخل إليه الطالب، فقد أشارت بعض الأدبيات على أن معاملات الثبات في الأدوات والبيانات التي تستخدم لغايات اتخاذ القرار يجب أن لا تقل عن ٠,٧ (Feldt & Bernnan, 1989) ، في حين أشارت دراسات أخرى على أنه يجب أن لا تقل عن ٠,٨ (Frary, 2000; Runer & Schafer, 2001) لأغراض تنبؤية، فقد أشارت بعض الدراسات مثل دراسة نوبل (Noble, 1991) ودراسة صوير (Sawer, 1989) إلى أن ضعف الثبات في العلامات الجامعية كان من الأسباب وراء تدني القدرة التنوبية لامتحانات القبول في الجامعات الأمريكية. علماً بأن معاملات الثبات بشكل عام في هذه الدراسة أعلى مما أشارت إليه نتائج الدراسات المماثلة في كثير من الجامعات (الشايب،٢٠٠٧؛ الأحمد، ٢٠١٠؛ العمري والعكور، ٢٠١٤)، فهي من الناحية النسبية مرتفعة، إلا أنها لا تتلاءم مع رؤية ورسالة جامعة تبوك القائمة على التميز والتي تتطلب رفع سقف معاييرها، ومن ضمنها دقة العلامات الجامعية.

وأظهرت نتائج هذه الدراسة وجود علاقة بين نوع المقرر ومستوى الثبات؛ إذ تختلف أعداد ونسب المقررات في كل مستوى من مستويات الثبات باختلاف المقرر، ونظرا لكثرة العوامل التي يمكن أن تتحكم بدقة درجات الطلبة، فيتوقع لأن تدفع نتائج هذه الدراسة الى دراسات أخرى تحليلية وتفسيرية على مستوى المقرر الواحد في ضوء تعدد الشعب وتعدد المدرسين، وانعكاس ذلك على جودة الاختبارات وأدوات التقويم الأخرى، ومستوى التنسيق بين المدرسين ، والتعريفات الاجرائية لتوحيد الاختبارات، وانعكاس ذلك على المقاييس الاحصائية لتوزيع الدرجات بشكل عام والانحراف المعياري بشكل خاص. وقد يكون لفلسفة عضو هيئة التدريس في قضية وضع العلامات وخصوصاً المتدنية منها دوراً مهماً في التأثير على تقدير الثبات، فرغم وجود توزيع افتراضي لنظام الدرجات والتي تضعه الجامعة، ويبدأ بالعلامة صفر وينتهي بالدرجة مائة ووجود درجة قطع للنجاح في المقرر إلا أن ما ينظر إليه المدرس يختلف مع ذلك، فقد يحاول أن يصل بالدرجات المتدنية أن تكون قريبة من درجة القطع أو علامة النجاح في المقرر، وبالتالي يخلق هذا ما يسمى بظاهرة ضيق المدى عند الطلبة restriction of range، والتي تؤثر بدورها سلباً في نتائج تقديرا الثبات.

وما يقال عن تأثير متغير نوع المقرر يمكن أن يقال عن متغير الشطر، ليس بسبب الجنس النوع الاجتماعي) للطالب، وانما قد يكون هناك عوامل غير مباشرة متعلقة بهذا المتغير، فقد يكون اختلاف جنس عضو هيئة التدريس وليس جنس الطالب، هو الذي أدت الى وجود علاقة بين مستوى الثبات والشطر للمقررات المختلفة، حيث ظهرت نسب أعلى للمقررات في المستوى المرتفع من الثبات في شطر الطالبات عنها في شطر الطلاب، وربما يكون اختلاف توزيعات العلامات تبعاً للشطر أحد العوامل المسببة لذلك، وربما يعود الى اختلاف في جودة الاختبارات، ولا يوجد تفسير واحد لهذا الفرق نظراً لكثرة العوامل المحددة لقيمة الثبات.

وبصرف النظر عن نوع المقرر أو الشطر، وما يتعلق بهما من عوامل ديموغرافية، فإن ما يتحكم بالثبات وفق التعريف الاحصائي للثبات هو تباين الدرجات أو الانحراف المعياري الذي يحرص عضو هيئة التدريس الى النظر اليه الى جانب المتوسط الحسابي. فربما

يكتشف عضو هيئة التدريس بعد إجراء اختباره أن جميع درجات الطلبة متقاربة أو جميعهم حصلوا على درجات عالية، وهذا حتماً سيخفض قيمة الانحراف المعياري لدرجات طلاب شعبته، علما بأن معاملات الثبات التي تم حسابها لأغراض المقارنة وفق متغير المقرر والشطر اعتمدت الأوزان النسبية لأدوات التقويم التي تحددها الجامعة في التعليمات لائحة الدراسة والاختبارات) بصرف النظر عن الانحرافات المعيارية لتلك الدرجات. ولذا جاءت الاضافة النوعية في هذه الدراسة للنظر من زاوية أخرى الى معادلة راجو. حيث تم تقدير الثبات باستخدام الانحرافات المعيارية كأوزان نسبية، وهي متوفرة عادة في التحليلات البسيطة للدرجات، وبالتالي تقع ضمن إمكانات عضو هيئة التدريسFeasible). وأظهرت النتائج بشكل عام تقديرات أعلى إحصائيا لمعاملات الثبات باستخدام هذه الأوزان مقارنة بالأوزان القائمة على التعليمات وذلك بفص الفروق باستخدام اختبار "ت" للعينات المرتبطة (Paired Sample t-test)، والتي أظهرت وجود فروق دالة إحصائياً لصالح طريقة الأوزان للانحرافات المعيارية في غالبية المقررات، ومن المتوقع أن تكون الدقة في تقدير الثبات أعلى لأن كل اختبار أخذ دوره الحقيقي في معادلة الثبات مع الافتراض بأن للتباينات المشتركة دورا ضعيفا في التأثير على دقة تقدير معاملات الثبات الذي تتعامل معها معادلة فلدت –جيلمر، ولفحص هذا الافتراض، لمعرفة مدى امكانية مقاومة أو تجاوز هذه المعادلة لانتهاك هذا الافتراض كما اتضح في مشكلة الدراسة وأسئلتها وإجراءاتها الاحصائية علي عينة مختارة من الشعب الأصلية، فلم تظهر المقارنات الاحصائية فروقا ذات دلالة إحصائية في متوسطات الثبات بين طريقة تقدير الثبات بالطرق الثلاث المشار اليها، وقد كشفت النتائج عن صحة هذا الافتراض في هذه الدراسة، الا أنها غير قابلة للتعميم، ويحتاج ذلك الى المزيد من الدراسات الاحصائية، الا أن العمليات الاحصائية التي لا تقع ضمن الامكانات المتاحة لأعضاء هيئة التدريس في غياب البرمجيات، يمكن التوصية باعتماد معادة راجو المعدلة والقائمة على الانحرافات المعيارية للدرجات في تقدير الثبات.

الاستنتاجات والتوصيات

١. رغم ما أظهرته هذه الدراسة في أن معاملات الثبات في مقررات جامعة تبوك أعلى مما تشير إليه الدراسات المماثلة في كثير من الجامعات، إلا أن الخطة الاستراتيجية للجامعة القائمة على التميز تتطلب رفع سقف معاييرها، ومن ضمنها دقة العلامات الجامعية، لذلك توصي الدراسة بضرورة الحدّ من العوامل التي تؤثر سلباً على تقدير معاملات الثبات في المقررات، ويتطلب ذلك تدريب أعضاء هيئة التدريس على إعداد الاختبارات في دورات وورش تدريبية موجهة لهذا الغرض.

٢. تبين في هذه الدراسة وجود اختلاف في تقدير معاملات الثبات من مقرر إلى أخر، وكذلك اختلاف في تقدير معاملات الثبات للمقررات من شطر إلى أخر (طلاب، طالبات)، ولذلك تأتي أهمية دراسة عدالة الاختبارات وتحيزها المحتمل، ومدى انعكاس ذلك على تقديرات الثبات للمقررات والشعب في الشطرين.

٣. تتعامل معادلة راجو (Raju) في الأصل مع الأطوال (عدد الفقرات) كأوزان نسبية، وجاءت هذه الدراسة لإلقاء الضوء على أن هذا يقع ضمن محددات استخدامها، وأن هناك استراتيجيات أخرى في التعامل مع الأوزان النسبية، وهو الأوزان في التعليمات مرة، والانحرافات المعيارية مرة أخرى مع افتراض مقامة المعادلة للتباينات المشتركة التي تعاملت معها معادلة فيلدت جيلمر، وتشجع النتائج، وفقا للمعايير الاحصائية والمنطقية، أعضاء هيئة التدريس لاستخدام هذه المعادلة في تقدير ثبات نتائج اختباراتهم. وبناء عليه توصي الدراسة بشرح هذه

الطريقة لأعضاء هيئة التدريس في ورش تدريبية موجهة لتعزيز الخصائص السيكومترية للاختبارات التي يعدها أعضاء هيئة التدريس في الجامعات، لسهولة توفير متطلبات هذه المعادلة مقابل ما تتطلبه معادلة فيلدت – جيلمر من حسابات قائمة على تجهيز مصفوفة التينات والتباينات المشتركة كخطوة أولى للمزيد من الحسابات لتحديد الأوزان النسبية، حيث يتوقع أن لا تكون بمقدور الكثير من أعضاء هيئة التدريس توفيرها ضمن الإمكانات المتاحة؛ وفي الوقت نفسه توصي الدراسة بتجهيز دليل لعضو هيئة التدريس في هذا الإطار مرفقا ببرنامج software يسهّل إجراءات التحليل في أي من الطرق الثلاث، وأي طريقة أخرى في المواقف والمتغيرات المختلفة التي تشير إليها معادلات تقدير الثبات.

المراجع
References

أبو جراد، حمدي (٢٠٠٩). دراسة سيكو مترية للعلامات الجامعية في عينة من المقررات في جامعة القدس المفتوحة. **المجلة المصرية للدراسات النفسية، (٦٦)**، ٢٣ -٥٥.

الأحمد، محمد (٢٠١٠). **تقديرات الثبات للعلامات الجامعية في جامعة جدار باستخدام معادلة راجو للاختبارات المتشاكلة.** رسالة ماجستير غير منشورة، جامعة جدارا، الأردن.

السوالمة، يوسف (١٩٩٥). تقدير الثبات للعلامات الجامعية في عينة من المساقات في جامعة اليرموك خلال الفصل الدراسي الثاني لعام ١٩٩٣/١٩٩٢ م. **مجلة مركز البحوث، جامعة قطر (٧)**، ٧١ -٨٩.

السوالمة، يوسف (٢٠٠١). أثر تحليل نتائج الطلبة في المساقات الجامعية على ثبات الاتساق الداخلي للدرجات الجامعية. **المجلة التربوية**، ١٥ (٨٥)، ١٥١ -١٧٥.

الشايب، عبد الحافظ (٢٠٠٧). تقدير الثبات في عينة من المواد في جامعة آل البيت. **مجلة جامعة دمشق للعلوم التربوية، (٢٣٤)** ٢٥٥ -٢٧١.

عليان، ريم (٢٠٠٣). **تقدير الثبات للعلامات المدرسية في مدينة إربد الكبرى في الفصل الدراسي الأول ٢٠٠٣/٢٠٠٢م.** رسالة ماجستير غير منشورة، جامعة اليرموك، اربد، الأردن.

العمري، حسان والعكور، معتصم (٢٠١٤). أثر حجم الشعبة الدراسية على تقديرات الثبات للعلامات الجامعية. **مجلة العلوم التربوية**، جامعة الكويت، ١٥ (٣)، ٤٤ -٦٧.

المسند، شيخة (١٩٩٢). دراسة تحليلية لتوزيع تقديرات الطلاب في جامعة قطر في الفصل الدراسي خريف ١٩٩٠م. **حولية كلية التربية**، جامعة قطر (٩)، ١٢٥ -١٥٣.

Blight, D. (1988). *Higher Education*. London: Cassel Educational Limted.

Division of Instructional Innovation and Assessment (DIIA) (2003). *Test Item Analysis & Decision Making*. Retrieved 20-9-2016 from http://www.utexas.edu/academic/dia.

Etaugh, A., Etaugh, C., & Hurd, D. (1972). Reliability of college grades and grade point average: some implications for prediction of academic performance. *Educational and psychological Measurement,* 32, 1045-1050.

Feldt, L. & Bernnan, R. (1989). Reliability. In R. Linn. *Educational measurement*3rd ed.).New York: American Council on Education.

Feldt, L. & Brennan, R. (1989). Reliability. In Linn, R.Ed.), Educational Measurement, New York: Macmillan Publishing Company.

Frary, R. (2000). Reliability of test scores. Retrieved 2-7-2016 from: http://www.testscoring.Vt.Edu/memo08.htm1.

Gronlund, N. & Robert, L. (1990). *Measurement and evaluation on teaching. (6th Edition).* New York: Macmillan.

Hathocat, J. (2013). Validity semantics in educational and psychological assessment. *Practical assessment, Research and Evaluation 18 (9),* 1-14.

Iukke, A. (2014). Validity and reliability of teacher-made tests: Case study of year 11 Physics in Nyahururu District of Kenya. *African Educational Research Journal*, 2(2), 61-71.

Kingston, N. & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28-37.

Nobel, J. (1991). Predicting College course grades using ACT *Assessment scores and high school course work information (ACT Research Report).* Iowa: ACT.

Notar, C., Zuelke, D., Wilson, J., & Yunker, B. (2004).The table of specifications: Insuring accountability in teacher mad tests. *Journal of Instructional Psychology*, 31, 115-129.

Parr, A. & Bauer, W. (2006). *Teacher made test reliability: a comparison of test scores and student study habits from Friday to Monday in a high school biology class in Monroe* County O*hio. Master Thesis. Graduate School of Marietta College. Retrieved april,12,1972 From*

https://etd.ohiolink.edu/rws_etd/document/get/marietta1142864088/inline.

Raju, N. (1977). A generalization of coefficient alpha. *Psychometrika,*42, 549-565.

Rudner, L. & Schafer, W. (2001). *Assessment and evaluation- College Park*. (Reports-Research (1129).U.S: University of Maryland: Clearing house on Assessment and Evaluation College Park, MD. (ERIC Document Reproduction service No. ED 458213).

Sawer, R. (1989).*Validating the use of ACT assessment scores and high school course work information (ACT Research Report).* Iowa: ACT.

Shoenfeldt, l. & Brush, D. (1975). Patterns of college grades across curricular areas: Some implication for GPA as criterion. *American Educational Research Journal,* 12 , 313-321.

Thorndike, R. & Hagen, E. (1986). *Measurement and evaluation in psychology and education (4th Edition).* New York: Macmillan.

Wells, C. & Wollack, J. (2003). *An instructor's guide to understanding test reliability.* Testing & Evaluation Services, University of Wisconsin.

Worthen, B., Brog, W. & White, K. (1993). *Measurement and evaluation in the schools.* White Plains, NY: Longman.