

أثر تحليل نتائج الطلبة في المسابقات الجامعية على ثبات الاتساق الداخلي للدرجات الجامعية

د. يوسف محمد سوامه*

* قسم علم النفس - جامعة الإمارات العربية.

أثر تحليل نتائج الطلبة في المسابقات الجامعية على ثبات الاتساق الداخلي للدرجات الجامعية



ملخص

هدفت الدراسة إلى تحديد أثر تحليل نتائج الطلبة في المسابقات الجامعية على ثبات الاتساق الداخلي للدرجات الجامعية. وتكونت بيانات الدراسة من الدرجات النهائية والفرعية للطلبة في ١٠٨٧ شعبة لمسابقات جرى تدريسها في كلية العلوم الإنسانية والاجتماعية في جامعة الإمارات العربية المتحدة خلال العام الدراسي ١٩٩٨/٩٧. وقد رصدت درجات الطلبة في شعب الفصل الأول (٥٣٠ شعبة) قبل القيام بأول دراسة تحليلية لنتائج الطلبة، بينما رصدت درجات الطلبة في شعب الفصل الثاني (٥٥٧ شعبة) بعد إجراء الدراسة التحليلية لنتائج الطلبة والاستفادة منها في تحسين الممارسات التقويمية. ولتقدير ثبات الدرجات الجامعية في كل شعبة تم تطبيق معادلة راجو. وأشارت النتائج إلى أن الدرجات الجامعية في الفصلين تتمتع بمستوى مرتفع من الثبات، حيث يزيد معامل الثبات في ٥٨٪ من المسافات على ٠,٧٠، كما أنه توجد فروق ذات دلالة إحصائية في توزيعات الثبات بين الفصلين لصالح الفصل الثاني، حيث ارتفعت نسبة المسابقات ذات الثبات المرتفع نتيجة الاستفادة من تحليل نتائج الطلبة من ٥٣٪ في الفصل الأول إلى ٦٢٪ في الفصل الثاني.

مقدمة :

يشكل التقويم بعداً مهماً في العملية التعليمية، وتبرز أهميته في المحافظة على نوعية وجودة البرامج التعليمية التي تقدمها مؤسسات التعليم المختلفة، ومن خلاله يتم توثيق مدى أهمية تلك البرامج في إحداث التغيرات السلوكية المرغوب فيها لدى الطلبة؛ (الطويل، ١٩٨٦). وبغياب التقويم لا يمكن لعملية التعليم أن تكتمل أو تنجح في إحداث التغيرات المنشودة. ويحول قصور عملية التقويم وضعها دون معرفة المؤسسات التعليمية بمدى نجاحها في تحقيق أهدافها؛ (أبو لبدة، ١٩٨٧؛ عجاوي وخضر، ١٩٨٧).

ويهدف التعليم الجامعي إلى إعداد الأفراد للحياة. ومن أجل ذلك، ينبغي أن تركز الجامعات على تعزيز مدخلاتها وعملياتها للوصول إلى المستويات العليا في التحصيل الأكاديمي والإعداد المهني لطلبتها؛ (المطوع، ١٩٨٨)، وتهتم بتقويم طلبتها ومخرجات التعليم فيها. ويعتبر الاهتمام بنظم التقويم وتطويرها من الأولويات المرتبطة ارتباطاً مباشراً بموثوقية نتائج التقويم في التطوير التربوي؛ (بوبطانه، ١٩٨٦). كما ينظر إلى تغيير أنظمة وأساليب التقويم باعتباره من أسرع الطرق لتغيير أنماط التعلم لدى الطلبة؛ (Elton and Laurillard, 1979).

ويعتبر التحصيل العلمي محوراً مهماً من محاور الاهتمام الرئيسية في التعليم الجامعي. ويتطلب الارتقاء بمستوى التحصيل ارتباطاً وثيقاً بالتقويم المستمر والشامل على مستوى الفرد والمؤسسة؛ (العريض المشار إليه في فرحات، ١٩٩٧). ويشكل رصد الدرجات وتحديد التقديرات غرضاً من أغراض التقويم؛ إلا أن رصد الدرجات ليس هدفاً بحد ذاته، وإنما وسيلة لأغراض متعددة. ومن الضروري أن يكون للدرجات معنى، وأن تكون ذات مدلول واضح؛ (عودة، ١٩٩٣).

وتشكل الدرجات وما ينبثق عنها من تقديرات أساساً مهماً للكثير من الإجراءات والقرارات الهامة التي ترتبط بوضع الفرد وتؤثر عليه. فأهلية الفرد للاستمرار بالدراسة، أو القبول في برنامج معين، أو الحصول على بعثة دراسية أو وظيفة معينة تتقرر بالمستوى الأكاديمي الذي يحققه ممثلاً في الدرجات أو التقديرات التي يحصل عليها.

وتنظر المؤسسات التعليمية إلى الدرجات والتقديرات التي ترتبط بها باعتبارها الوسيلة الرئيسية المتاحة لتحديد مستوى كفاءة الطالب ومقدرته على الوفاء بالمستويات الأكاديمية التي تتطلبها تلك المؤسسات، وللتنبؤ بمقدرته على النجاح في دراسته المستقبلية.

وتؤدي الدرجات وظائف فعلية ترتبط بحاجات الطلبة وأولياء الأمور والمدرسين وأصحاب العمل. فهي تلعب دوراً مهماً في تكوين الطالب صورة عن ذاته، وستبقى من أفضل عوامل التنبؤ بالأداء اللاحق؛ (ثورندايك وهيجن،

(١٩٨٩). ولهذا فإن من الضروري أن تمتاز الدرجات في أية مؤسسة تعليمية بقدر كبير من الصدق والثبات وأن تكون أساليب تقديرها سليمة ولا تخضع لعوامل شخصية. والنظام الجيد للدرجات هو ذلك النظام الذي يسمح بتقدير تحصيل الطالب للأهداف المحددة دون أن تتأثر الدرجات بالانطباعات الشخصية وعوامل الجهد والمثابرة والتحسين؛ (عودة، ١٩٩٣؛ Frisbie, 1983).

وينظر كثير من المربين إلى عملية رصد الدرجات باعتبارها تلخيصاً للمعلومات المتوافرة عن مدى تقدم الفرد نحو تحقيق الأهداف على صورة رموز أو أرقام للاستدلال على مستوى الأداء وجودته؛ (Payne, 1974). كما أن هناك اتفاقاً على أنه لا بد أن يكون التحصيل هو العامل الأساسي في تقدير الدرجات من أجل أن تكون الدرجات ترجمة دقيقة للتحصيل الحقيقي؛ (Merwin, 1989; Stiggin, Frisbie and Griswold, 198).

ومن الناحية العملية، فإن الدرجات تعبر عن نظم القيم لدى الأفراد بقدر ما هي تقارير حيادية عن سلوك المتعلم؛ فقد تتأثر عملية رصد الدرجات بأحكام المدرس الشخصية وبالسياسة التقويمية المتبعة في المؤسسة التعليمية. ويذكر سوامه (٢٠٠٠) أن أعضاء هيئة التدريس يعتمدون على التحصيل وعلى عوامل أخرى غير التحصيل في تقدير الدرجات الجامعية، وأن معنى الدرجة يرتبط ارتباطاً وثيقاً بعمل الطالب، فالدرجة أجر يدفعونه للطالب نظير ما يؤديه من أعمال، وأكثر ما يهمهم في تقدير الدرجات الاحتكام إلى قيم العدل والإنصاف وقيم الرحمة والرفقة. كما أن أستاذ الجامعة الذي يشبع غروره بشعار المحافظة على المستوى، سيكون راضياً عن ارتفاع نسبة الرسوب وانخفاض عدد الذين يحصلون على درجات عالية عنده؛ (ثورندايك وهيجن، ١٩٨٩).

وتتضمن العوامل غير التحصيلية التي تراعى في عملية رصد الدرجات الجهد المبذول والتحسين والمشاركة والسلوك الصفي والمداومة على الحضور والسمات الشخصية وحجم العمل المنجز والسلوك الاجتماعي؛ (Barnes, 1985; Noble, 1991). وبالرغم من أهمية تشجيع التقدم وبذل الجهد والمثابرة، إلا أنه

يجب أن لا يكون هذا على حساب دقة الدرجات ومعناها. فالدرجة الكلية في المقرر الدراسي يجب أن تبقى مقياساً نقياً للتحصيل، وأن ترتبط بمجموعة معرفة بدقة من الأهداف التدريسية (Terwilliger, 1989)، وأن تعكس مستوى ما يتحقق من نتائج تعليمية؛ (Manke and Loyed, 1991). ويرى جرونلاند (Gronlund, 1985) أنه يمكن استخدام مؤشرات منفصلة للتحسن وبذل الجهد والخصائص الشخصية الأخرى من أجل استثارة دافعية المتعلم للحصول على أداء أفضل.

وفيما يتعلق بالسياسة التقييمية للمؤسسات التعليمية، نجد أن بعض المؤسسات تضع قيوداً على نسبة الرسوب أو نسبة من يحصلون على تقدير «ممتاز»، مما يضطر المدرس إلى تحويل الدرجات بالشكل الذي يجعل توزيع الدرجات ضمن المدى الذي تحدده المؤسسة التعليمية. وقد يثير تحويل الدرجات بهذه الصورة التساؤلات حول مدى تمثيل الدرجة لكفاية الفرد، ويزداد التشكيك بأهمية الدرجات عندما تنقصها الدقة والثبات وتكون قدرتها التنبؤية بالأداء المستقبلي منخفضة؛ (عودة، ١٩٩٣).

ويتوافر في التعليم الجامعي أساليب متعددة لجمع البيانات عن مقدار ما يحققه الطالب من أهداف المساقات الجامعية (المقررات الدراسية الجامعية). وتتمثل هذه الأساليب في أنواع من التدريبات والواحدات، وما قد يُطلب من أبحاث وتقارير، وما يُطبق من اختبارات قصيرة وامتحانات. وتوفر هذه الأساليب مؤشرات وكواشف منفصلة يجب أن تدمج بطريقة مناسبة للحصول على الدرجة النهائية (الكلية) في المساق. ويتم ذلك بإعطاء كل منها وزناً مناسباً من الدرجة الكلية. ومن الضروري أن يعتمد الوزن المناسب لنوع البيانات المتحققة من نشاط معين أو تدريب أو تقرير أو اختبار على صدق البيانات وثباتها؛ (ثورندايك وهيجن، ١٩٨٩). ويعالج إعطاء وزن لأعمال السعي في عملية تحديد الدرجات الكثير من الظواهر التي تشكو منها المؤسسات التعليمية وفي مقدمتها مشكلات السلوك والمواظبة؛ (غوني، ١٩٩٢). والملاحظ أن أساليب تطبيق نظام أعمال السعي لا تقوم على أساس علمي سليم؛ (سرحان،

(١٩٨١). ويؤدي هذا الاستخدام غير السليم إلى خلل في عملية التقويم، ومن مظاهر هذا الخلل أن تتجه تقديرات الطلبة إلى الأعلى بصورة عامة؛ (عجاوي وخضر، ١٩٨٧).

وتعاني الكثير من المؤسسات التعليمية من ظاهرة ارتفاع المعدلات. وقد بدأت هذه الظاهرة في الجامعات والكليات الأمريكية في منتصف الستينات، حيث لاحظ سسلو (Suslow, 1976) ارتفاع المتوسط الحسابي للمعدلات التراكمية في (١٥) جامعة أمريكية رئيسية من ٢,٤٩ في سنة ١٩٦٣ إلى ٢,٩٤ في سنة ١٩٧٤. وفي سنة ١٩٧٤، كان نصيب التقديرات في فئتي «أ» و«ب» ٧١٪ من مجمل التقديرات.

ويتوقع أن يؤثر التضخم في التقديرات بشكل سلبي على صدقها وثباتها (Davidson, 1975). ويعني التضخم في التقديرات ارتفاع نسبة التقديرات المرتفعة بصورة واضحة على حساب نسبة التقديرات المنخفضة والمتوسطة، مما يجعل التمييز بين الطالب المتفوق والطالب المتوسط مهمة بالغة الصعوبة. وبالفعل فقد وصفت نظم تقدير الدرجات الجامعية بالفساد وعدم الثبات (Bligh, 1988)، وبأنها تفتقر إلى الموضوعية والشمول (المسند، ١٩٩٢)، وبدأ المدرسون التخلي عن دورهم التقويمي وأصبحوا يطبقون معايير متدنية للأداء المقبول؛ (White, 1975).

ويُعَيّن التقدير في المساق الجامعي من خلال تقدير التحصيل للفرد على تدرج متعدد الفئات، تبدأ نقاطه من هـ (الرسوب) وتنتهي عند أ (الامتياز). ويؤكد الاختصاصيون في القياس النفسي والتربوي على وجود علاقة طردية بين ثبات مقياس التقدير وعدد فئات (نقاط) التدرج (Masters, 1974). ومن المحتمل أن يؤدي التضخم في الدرجات إلى انخفاض عدد نقاط التدرج، وينتج عن ذلك انخفاض قيمة الثبات. ويهدف التغلب على هذه المشكلة قامت الكثير من الجامعات بزيادة عدد نقاط تدرج التقديرات الجامعية من خلال استخدام التقديرات الموجبة (أ⁺، ب⁺، ج⁺، د⁺، هـ⁺).

وتطبق جامعة الإمارات تدرجاً ثمانياً للتقديرات (أ⁺، ب⁺، ج⁺، ج⁺، د⁺، د⁺، هـ). بدلاً من التدرج الخماسي (أ⁺، ب⁺، ج⁺، د⁺، هـ) منذ عام

١٩٨٩. وتتحدد درجات الطالب في الجامعة وفق النظام المطلق للدرجات حيث تعتمد على مستواه في إتيقان المحتوى، والدرجة النهائية في المساق هي محصلة مجموع (١) درجة أعمال السعي، و(٢) درجة امتحان نصف الفصل، و(٣) درجة امتحان نهاية الفصل. ويتم تقييم الدرجات النهائية في المساق وتحديد التقديرات على النحو المبين في الجدول رقم (١)، وذلك بتصنيفها إلى ثمانية مستويات.

وقد دأبت كلية العلوم الإنسانية والاجتماعية في الجامعة اعتباراً من الفصل الثاني من العام الدراسي ١٩٩٨ / ٩٧ على تبني سياسة تحليل النتائج ودراساتها من أجل: (أ) تحديد النمط العام لتوزيع التقديرات في الكلية، وفي كل قسم من أقسامها العلمية، ولدى كل عضو من أعضاء هيئة التدريس فيها، (٢) بحث العلاقة بين توزيع التقديرات ومجموعة من المتغيرات المرتبطة بالمساق (طبيعته، مستواه، ولغة تدريسه) والمدرس (رتبته، جنسه، طبيعة عقده، عدد سنوات عمله في الجامعة، وتقييم الطلبة له) وجنس الطالب، (٣) ودراسة العلاقة بين مكونات الدرجة النهائية في المساق.

جدول رقم (١)

تقييم الدرجة النهائية في جامعة الإمارات العربية المتحدة

الدرجة النهائية	التقدير	الرمز	النقاط
٩٠ - ١٠٠	ممتاز	أ	٤,٠
٨٥ - أقل من ٩٠	جيد جداً مرتفع	ب+	٣,٥
٨٠ - أقل من ٨٥	جيد جداً	ب	٣,٠
٧٥ - أقل من ٨٠	جيد مرتفع	ج+	٢,٥
٧٠ - أقل من ٧٥	جيد	ج	٢,٠
٦٥ - أقل من ٧٠	مقبول مرتفع	د +	١,٥
٦٠ - أقل من ٦٥	مقبول	د	١,٠
أقل من ٦٠	راسب	هـ	صفر

وتتميز الدراسات التحليلية للنتائج التي تقوم بها الكلية عن الدراسات السابقة التي اهتمت بتوزيع التقديرات على مستوى الجامعات الخليجية (أبو علام والصراف، ١٩٨٥ أ، ١٩٨٥ ب؛ عجاوي وخضر، ١٩٨٧؛ المسند، ١٩٩٢) بأنها: (١) تعبر عن سياسة ثابتة تتبناها الكلية حيث يتم التحليل للنتائج في كل فصل دراسي، ويكتب تقرير شامل حول نتائج التحليل ويرسل إلى الأقسام العلمية لمناقشته والاستفادة منه في تحسين الممارسات التقويمية، (٢) تتصف بالشمول حيث تناول عدداً كبيراً من المتغيرات المرتبطة بالطلاب والمدرس والمساق وتأخذ بالاعتبار جميع النتائج في الكلية.

وقد أجريت أول دراسة شاملة لتحليل النتائج في الفصل الثاني من العام الدراسي ١٩٩٨/٩٧، حيث حللت نتائج الفصل الأول من ذلك العام (شهادة وآخرون، ١٩٩٨). وبينت نتائج التحليل وجود فروق ذات دلالة إحصائية بين توزيعات التقديرات الجامعية على مستوى الكلية تعزى إلى اختلاف كل من القسم العلمي، ولغة تدريس المساق، ومستواه، وطبيعته في الخطة الدراسية، والصفة الوظيفية (طبيعة العقد) للمدرس، ورتبته الأكاديمية، وجنسه، وعدد سنوات عمله في جامعة الإمارات، ودرجة تقييم الطلبة للتدريس، وجنس الطلبة. وكانت هناك نتائج مماثلة على مستوى الأقسام العلمية في الكلية.

في ضوء هذه النتائج يمكن تفسير جانب كبير من الاختلافات بين توزيعات التقديرات الجامعية لدى أعضاء هيئة التدريس سواء على مستوى الكلية أو الأقسام العلمية فيها. لكنه في الوقت نفسه، يصعب توظيفها في تفسير الاختلافات الكبيرة بين توزيعات التقديرات في المساقات متعددة الشعب حيث تضبط معظم المتغيرات المرتبطة بالمساق وتكون مجموعات الطلبة في الشعب المتعددة متكافئة إلى حد ما. في هذه الحالة، وعندما يختلف توزيع الدرجات لدى مدرس بصورة واضحة عن توزيعات الآخرين، فإنه يجب عليه إعادة النظر في ممارساته التقويمية ومراجعتها للتأكد من أنها معقولة ومبررة ويمكن الدفاع عنها؛ (Frisbie, 1983).

ومن الناحية العملية، أرسلت نتائج التحليل الخاصة بكل قسم علمي إلى ذلك القسم وطلب منه دراسة تلك النتائج والاستفادة منها في تحسين أساليب تقويم الطلبة. وقامت الأقسام فعلاً بدراسة تلك النتائج ووظفتها في تعديل سياساتها التقويمية وبالأخص في حالات التطرف التي تنطوي على مغالاة كبيرة في التشدد أو التساهل في معايير التقويم وإجراءاته. كما قام أعضاء هيئة التدريس الذين بينت الدراسة وجود تطرف في توزيع التقديرات عندهم بمراجعة الممارسات التقويمية التي يتبعونها في تقويم طلبتهم للاطمئنان على سلامتها ولتطويرها بحيث تفرز درجات تعبر فعلاً عن القدرات الحقيقية للطلبة.

أشارت نتائج الدراسة التحليلية لنتائج الفصل الثاني من العام الدراسي ١٩٩٨/٩٧ (شهادة وآخرون، ١٩٩٨ ب) إلى اقتراب توزيعات التقديرات في معظم الأقسام العلمية من التوزيع المعتدل، ونقصان حالات التطرف، وازدياد قوة العلاقة بين مكونات الدرجة النهائية عما كانت عليه في الفصل الأول. وفي الوقت الذي قد ترتبط فيه هذه التغيرات الإيجابية باختلاف طبيعة المساقات المطروحة في الفصل الثاني عن المساقات التي كانت مطروحة في الفصل الأول، وبمتغيرات أخرى، إلا أن جزءاً من التحسن الذي طرأ على توزيع التقديرات يرتبط بمدى الاستفادة من نتائج الدراسة التحليلية لنتائج الفصل الأول.

وهكذا فإن تحليل النتائج يشكل خطوة هامة نحو كشف بعض جوانب الخلل في الممارسات التقويمية مما يفتح المجال لمراجعتها وتطويرها بما يسهم في تحسين العملية التعليمية. وتسعى كل دراسة تحليلية جديدة لنتائج الطلبة إلى معرفة مدى الاستفادة من الدراسات التحليلية التي تسبقها في تحسين الممارسات التقويمية بدلالة التحسن في شكل توزيع التقديرات الجامعية. أما الدراسة الحالية فتهتم بمعرفة مدى الاستفادة من تحليل نتائج الطلبة في تحسين الممارسات التقويمية بدلالة التحسن في ثبات الاتساق الداخلي للدرجات الجامعية. ويشير الاتساق الداخلي للدرجات الجامعية إلى الدرجة التي تشترك بها أساليب التقويم

المعتمدة في قياس تحصيل الفرد، وتعتمد قيمته على درجة اتساق أداء الفرد في هذه الأساليب. وتقدر قيمة ثبات الاتساق الداخلي في الحالية من خلال تطبيق معادلة راجو؛ (Raju, 1977).

وتوجد طرق مختلفة لتجميع الدرجات من أجل تحديد التقدير النهائي، ويؤكد سميث (Smith, 1979) على أن الطريقة الأفضل هي الأكثر ثباتاً. وتأخذ التوزيعات للدرجات أشكالاً متباينة في شدة التوائها وتفلطحها (Skewness & Kurtosis) في المسافات المختلفة، فالدرجات هي التي تفرض شكل التوزيع. وعليه، فإن دقة الدرجات وثباتها أكثر أهمية في تقويم الممارسات التقويمية من شكل التوزيع الذي تأخذه. لهذا يحرص الباحث في الدراسة الحالية على معرفة ما تتمتع به الدرجات الجامعية من ثبات، وبيان مدى التحسن في الثبات كنتيجة لتبني سياسة تحليل النتائج. ويعتبر ثبات الدرجات الجامعية من المحكات الأساسية لتقييم الممارسات التقويمية المستخدمة في رصدها. وينبغي أن يؤدي تطوير الممارسات التقويمية إلى زيادة قيم الثبات للدرجات الجامعية.

إن الاستنتاجات المتعلقة بتدني الثبات للدرجات الجامعية ليس لها ما يبررها من ناحية علمية. وقد استندت هذه الاستنتاجات على وجود اختلافات في الممارسات التقويمية ولم تستند على أسس الحسابات المباشرة لمؤشرات الثبات (Etaugh, Etaugh and Hurd, 1972). فالمحاولات المبذولة لحساب الثبات للدرجات الجامعية قليلة جداً، ويعود ذلك إلى أن تقدير ثبات الدرجات الجامعية ليس بسهولة تقدير الثبات في الاختبارات، وهو يشكل مشكلة سيكومترية معقدة (Sawyer, 1989). فعادة، لا يعيد الطلبة دراسة ما درسوه من مساقات ولا توجد صور متكافئة للمساق نفسه، وعليه فإنه من غير الممكن استخدام طريقتي الإعادة والصور المتكافئة في تقدير ثبات الدرجات الجامعية. كما أنه لا يجوز استخدام طرق الاتساق الداخلي الشائعة (التجزئة النصفية، كرونباخ ألفا والحالات الخاصة المنبثقة عنها) لاعتبارات تتعلق بعدم تحقق شرط التكافؤ لمكونات الدرجة الجامعية.

وقد قام كلارك (Clark, 1950) بأول محاولة لتقدير ثبات الدرجات الجامعية، ويؤخذ عليه أنه استخدم تقديرات متحيزة للثبات (Etaugh et al., 1972). وفيما بعد، وفي دراستين منفصلتين (Etaugh et al., 1972; Schonfeldt and Brush, 1975) تم استخدام طريقة تحليل التباين لتقدير ثبات المعدلات الجامعية. ويؤخذ على تقديرات الثبات المستخلصة من هاتين الدراستين بأنها غير دقيقة بسبب عدم تحقق افتراض تكافؤ المساقات الجامعية. ويؤكد ديوك (Duke, 1983) أن المساقات الجامعية ليست متكافئة، وهذا يقلل من أهمية تحليل التباين في تقدير مؤشرات الثبات عندما يتعلق الأمر بالدرجات الجامعية. ويشير سوامه (١٩٩٥) إلى أنه من المناسب استخدام معادلة راجو لحساب مؤشرات دالة على ثبات الاتساق الداخلي للدرجات الجامعية. فالمعادلة تأخذ بالاعتبار عدم التكافؤ بين مكونات الدرجة الجامعية، ويمكن استخدامها حتى عندما تكون التباينات والأوزان لمكونات الدرجة النهائية غير متجانسة. وينطبق هذا الوضع على الدرجات الجامعية في جامعة الإمارات، حيث تتضمن كشوف النتائج الدرجات الفرعية لكل من أعمال السعي وامتحان المنتصف والامتحان النهائي. وبذلك يمكن الحصول على مؤشرات عن ثبات الدرجات الجامعية ممثلة بمعامل ثبات الاتساق الداخلي الذي هو محور الاهتمام في هذه الدراسة.

هدف الدراسة وأسئلتها:

تهدف هذه الدراسة إلى معرفة أثر تحليل نتائج الطلبة في المساقات الجامعية على ثبات الاتساق الداخلي للدرجات الجامعية في كلية العلوم الإنسانية والاجتماعية في جامعة الإمارات العربية المتحدة. وتحقيقاً لهذا الهدف، سيتم مقارنة توزيع قيم الثبات لدرجات الفصلين الأول والثاني من العام الدراسي ١٩٩٨/٩٧ حيث رصدت درجات الفصل الأول قبل تحليل النتائج بينما رصدت درجات الفصل الثاني بعد تحليل النتائج، وبالتحديد تسعى الدراسة إلى إجابة الأسئلة التالية:

- ١ - هل يؤثر تحليل النتائج على ثبات الدرجات الجامعية على المستوى العام للكلية؟
- ٢ - هل يؤثر تحليل النتائج على ثبات الدرجات الجامعية لكل من البنين والبنات في الكلية؟
- ٣ - هل يؤثر تحليل النتائج على ثبات الدرجات الجامعية في كل قسم من الأقسام العلمية في الكلية؟.

عينة الدراسة :

استخدمت لأغراض هذه الدراسة نتائج الطلبة في كلية العلوم الإنسانية والاجتماعية الخاصة بالفصلين الأول والثاني من العام الدراسي ١٩٩٨/٩٧. وبلغ عدد النتائج المستخدمة ٣٥١٤٦ نتيجة منها ١٧٥٤٤ نتيجة تتعلق بالفصل الأول. وقد تم الحصول على النتائج من كشوف الدرجات الخاصة بشعب المساقات المختلفة في الكلية. وقد اختيرت كل شعبة يزيد عدد طلبتها على خمسة. وبلغ عدد هذه الشعب ١٠٨٧ شعبة (٥٣٠ شعبة في الفصل الأول، ٥٥٧ شعبة في الفصل الثاني) منها ٢٤٠ شعبة للبنين (١١٩ في الفصل الأول، ١٢١ في الفصل الثاني) و٨٤٧ شعبة للبنات (٤١١ في الفصل الأول، ٤٣٦ في الفصل الثاني). وتم تقدير الثبات للدرجات الجامعية في كل شعبة، وبذلك توفر ١٠٨٧ معامل ثبات حيث شكلت قيم هذه المعاملات عينة هذه الدراسة.

المعالجة الإحصائية :

قدرت قيم ثبات الاتساق الداخلي للدرجات الجامعية في كل شعبة من خلال استخدام معادلة راجو التالية :

$$R = \left(\frac{1}{1 - \sum \lambda_i^2} \right) \left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2} \right)$$

حيث R : معامل ثبات الدرجات النهائية

λ_i : نسبة المكون ج في الدرجة النهائية

σ_i^2 : تباين الدرجات الفرعية للمكون i

σ_x^2 : تباين الدرجات النهائية.

فمثلاً: إذا كانت الدرجة النهائية تساوي مجموع درجات أعمال الفصل وامتحان المنتصف والامتحان النهائي، وكانت النسبة المئوية المخصصة للمكونات الثلاثة هي ٣٥٪ لأعمال الفصل و ٢٥٪ لامتحان المنتصف و ٤٠٪ لامتحان النهائي، وكانت قيمة التباين هي ٥٠ للدرجات النهائية و ٤ للدرجات أعمال الفصل و ٨ لدرجات امتحان المنتصف و ١٨ لدرجات الامتحان النهائي، فإن:

$$\lambda_1 = 0.35, \lambda_2 = 0.25, \lambda_3 = 0.40, \text{ وبالتالي } \sum \lambda_i^2 = 0.345, \sigma_1^2 = 4, \sigma_2^2 = 8, \sigma_3^2 = 18, \text{ وبالتالي } \sum \sigma_i^2 = 30, \sigma_x^2 = 50.$$

$$R = \left(\frac{1}{1-0.345} \right) \left(1 - \frac{30}{50} \right) = 0.61$$

وبهذا الأسلوب تم الحصول على ١٠٨٧ قيمة لمعامل الثبات في المسابقات المختلفة منها ٥٣٠ قيمة تخص مسابقات الفصل الأول التي رصدت درجاتها قبل تحليل النتائج، و ٥٧٧ قيمة تخص مسابقات الفصل الثاني حيث رصدت درجاتها بعد تحليل النتائج. وقد صنف قيم الثبات إلى ثلاثة مستويات: ثبات مرتفع حيث تزيد قيمة الثبات على ٠,٧٠، وثبات متوسط حيث تتراوح قيمة الثبات من ٠,٤٠ إلى ٠,٧٠، وثبات منخفض حيث تقل قيمة الثبات عن ٠,٤٠. ويأتي هذا التصنيف انسجاماً مع الاعتقاد السائد بأنه يلزم أن يكون معامل الثبات أكثر من ٠,٧٠ للأدوات والبيانات التي تستخدم لأغراض التقويم الفردي (Feldt & Brennan, 1989) ولإجابة أسئلة الدراسة جرت مقارنة التوزيعات التكرارية المتناظرة لقيم الثبات في الفصلين من خلال استخدام الاختبار الإحصائي مربع كاي.

النتائج:

يبين جدول رقم (٢) التوزيع التكراري لمعاملات الثبات في جميع المسابقات التي تناولتها الدراسة في كل من الفصلين. ويلاحظ في الجدول، أن عدد المسابقات في كل من الفصلين يزداد مع ارتفاع مستوى الثبات. وبصورة عامة، فإن النتائج في كل من الفصلين تتمتع بدرجة مقبولة من الثبات، حيث تقل نسبة المسابقات ذات الثبات المنخفض عن ١١٪ في الفصل الأول وعن ٨,٥٪ في الفصل الثاني.

جدول رقم (٢)
التوزيع التكراري لمعاملات الثبات لكل من الفصلين

مستوى الثبات		الفصل الأول		الفصل الثاني	
عدد المسابقات	النسبة المئوية	عدد المسابقات	النسبة المئوية	عدد المسابقات	النسبة المئوية
٥٧	١١	٤٧	٨	منخفض ($R > ٠,٤٠$)	
١٩١	٣٦	١٦٥	٣٠	متوسط ($٠,٤٠ > R > ٠,٧٠$)	
٢٨٢	٥٣	٣٤٥	٦٢	مرتفع ($R < ٠,٧٠$)	
٥٣٠	١٠٠	٥٥٧	١٠٠	المجموع	

ويلاحظ من مقارنة توزيعي الثبات، أن نسبة المسابقات ذات الثبات المرتفع قد ازدادت من ٥٣٪ في الفصل الأول إلى ٦٢٪ في الفصل الثاني، وأن نسبة المسابقات ذات الثبات المنخفض قد انخفضت من ١١٪ في الفصل الأول إلى ٨٪ في الفصل الثاني. ولمعرفة دلالة الاختلافات في توزيع قيم الثبات بين الفصلين، تم استخدام الاختبار الإحصائي مربع كاي بدرجات حرية (د.ح = ٢)، قيمة مربع كاي المحسوبة (٨,٥٣) ذات دلالة إحصائية عند مستوى $a = ٠,٠٥$. وتعني هذه النتيجة أن توزيع الثبات في الفصل الثاني أفضل منه في الفصل الأول.

ويبين جدول رقم (٣) توزيع المسابقات التي تناولتها الدراسة حسب متغيرات الجنس والفصل الدراسي ومستوى الثبات. وتشير الأعداد في داخل الجدول إلى أعداد المسابقات التي تتمتع بالمستوى المحدد للثبات. فمثلاً، من بين ١١٩ مساقاً مخصصة للذكور في الفصل الأول، يوجد ٢٨ مساقاً ذات ثبات منخفض و٤٢ مساقاً ذات ثبات متوسط و٤٩ مساقاً ذات ثبات مرتفع. ويلاحظ أن عدد المسابقات في فئة الثبات المرتفع أعلى من عدد المسابقات في كل من فئتي الثبات المتوسط والمنخفض، وينطبق هذا على الفصلين سواء بالنسبة للذكور أو الإناث، وبمقارنة توزيعي الثبات للذكور في الفصلين، يلاحظ ارتفاع نسبة المسابقات ذات الثبات المرتفع من ٤١٪ (٤٩ مساقاً من بين ١١٩) في الفصل

الأول إلى ٥٨٪ (٧٠ مساقاً من بين ١٢١) في الفصل الثاني، وانخفاض نسبة المسابقات ذات الثبات المنخفض من ٢٤٪ في الفصل الأول إلى ١٦٪ في الفصل الثاني، وهذه تغيرات إيجابية في توزيعات الثبات لمساقات الذكور. ولمعرفة الدلالة الإحصائية للاختلافات بين التوزيعين، جرى استخدام الاختبار الإحصائي مربع كاي. وتشير القيمة المحسوبة للاختبار الإحصائي مربع كاي (٦,٧٦) إلى أن للفرق بين التوزيعين دلالة إحصائية عند مستوى الدلالة ٠,٠٥.

وبالنسبة للإناث، ارتفعت نسبة المسابقات ذات الثبات المرتفع من ٥٧٪ (٢٣٣ مساقاً من بين ٤١١) في الفصل الأول إلى ٦٣٪ (٢٧٥ مساقاً من بين ٤٣٦) في الفصل الثاني، وانخفضت نسبة المسابقات ذات الثبات المنخفض من ٧٪ في الفصل الأول إلى ٦٪ في الفصل الثاني. وبالرغم من أن هذه التغيرات تشير بالاتجاه الصحيح، إلا أنها لم ترق إلى مستوى الدلالة الإحصائية ($\alpha = 0,05$). فالقيمة المحسوبة للاختبار الإحصائي مربع كاي (٣,٦٦) أقل من قيمته الحرجة (٥,٩٩) عند درجات الحرية (د. ح = ٢).

جدول رقم (٣)

توزيع المسابقات حسب متغيرات الجنس والفصل الدراسي ومستوى الثبات

قيمة الإحصائي مربع كاي	المجموع	مستوى الثبات			الفصل الدراسي	جنس الطالب
		مرتفع	متوسط	منخفض		
٦,٧٦ *	١١٩	٤٩	٤٢	٢٨	الأول	مسابقات الذكور
	١٢١	٧٠	٣٢	١٩	الثاني	
٣,٦٦	٤١١	٢٣٣	١٤٩	٢٩	الأول	مسابقات الإناث
	٤٣٦	٢٧٥	١٣٣	٢٨	الثاني	

* توجد دلالة إحصائية إحصائية عند مستوى الدلالة الإحصائية ٠,٠٥ (درجات الحرية = ٢).

جدول رقم (٤)

توزيع المساقات حسب القسم والفصل الدراسي ومستوى الثبات

القسم العلمي	الفصل الدراسي	مستوى الثبات			المجموع	قيمة الإحصائي مربع كاي
		منخفض	متوسط	مرتفع		
الدراسات الإسلامية	الأول	٤	٨	٤٢	٥٤	٠,٠٣
	الثاني	٤	٩	٤٣	٥٦	
اللغة العربية	الأول	٣	١٥	٤٦	٦٤	٠,٨٨
	الثاني	٢	١١	٤٩	٦٢	
التاريخ	الأول	١١	٢٩	١٤	٥٤	***١٦,٨
	الثاني	٤	١٤	٣٤	٥٢	
الجغرافيا	الأول	١٢	٢٥	٩	٤٦	**١١,٧٨
	الثاني	٧	١٨	٢٨	٥٣	
الفلسفة	الأول	٩	٥	٢٠	٣٤	*٦,٧٣
	الثاني	١	٨	٢٠	٢٩	
الاجتماع	الأول	٥	٢٨	١٩	٥٢	١,١٧
	الثاني	٣	٢٨	٢٥	٥٦	
اللغة الإنجليزية	الأول	٣	٩	٤٧	٥٩	٠,١٩
	الثاني	٢	٨	٤٦	٥٦	
الاتصال الجماهيري	الأول	٦	٢٢	٢٤	٥٢	١,٥
	الثاني	٦	١٤	٢٦	٤٦	
الخدمة الاجتماعية	الأول	٢	١٤	١٠	٢٦	٣,٧١
	الثاني	٧	٢١	١٨	٤٦	
العلوم السياسية	الأول	٤	١٢	١٨	٣٤	٠,٤٢
	الثاني	٣	١٠	٢٠	٣٣	
علم النفس	الأول	٧	٢١	٢٥	٥٣	٠,٤١
	الثاني	٨	٢٠	٣١	٥٩	
علوم الأسرة	الأول	٠	٣	٨	١١	٠,٦٤
	الثاني	٠	٤	٥	٩	

*** توجد دلالة عند ٠,٠٠١ ، ** توجد دلالة عند ٠,٠١ ، * توجد دلالة عند ٠,٠٥ (درجات الحرية = ٢) .

وفيما يتعلق بأثر تحليل النتائج على ثبات الدرجات الجامعية في الأقسام العلمية، فإنه يلاحظ من جدول رقم (٤) الذي يبين توزيعات الثبات في الفصلين الأول والثاني لكل قسم علمي وجود تغيرات إيجابية في توزيعات الثبات في معظم الأقسام العلمية. وتمثلت هذه التغيرات بانخفاض نسبة المسابقات ذات الثبات المنخفض أو بارتفاع نسبة المسابقات ذات الثبات المرتفع. وباستخدام الاختبار الإحصائي مربع كاي، تبين أن حجم هذه التغيرات لم يرق إلى مستوى الدلالة الإحصائية إلا في أقسام التاريخ والجغرافيا والفلسفة، ففي قسم التاريخ مثلاً، ارتفعت نسبة المسابقات ذات الثبات المرتفع من ٢٦٪ في الفصل الأول إلى ٦٥٪ في الفصل الثاني، وانخفضت نسبة المسابقات ذات الثبات المنخفض من ٢٠٪ في الفصل الأول إلى ٨٪ في الفصل الثاني. وقد كان لقيمة الإحصائي مربع كاي (١٦,٨) دلالة إحصائية عند مستوى الدلالة الإحصائية ٠,٠٠١.

مناقشة النتائج:

تشير بيانات الثبات المتوافرة من الدراسة إلى أن الثبات على مستوى الكلية في الفصل الثاني أفضل منه في الفصل الأول. وإذا ما افترضنا أن المجتمع الطلابي متكافئ إلى حد ما في الفصلين فإنه يمكن تفسير التحسن في ثبات الدرجات الجامعية باعتباره نتيجة لأحد أمرين أو كليهما:

١ - تبني سياسة تحليل نتائج الطلبة. فمن ناحية تُشعر هذه السياسة أعضاء هيئة التدريس بأن هناك متابعة وتقييم لممارساتهم التقويمية فيزداد اهتمامهم بها، ومن ناحية أخرى، تساعد على تعرف جوانب الخلل في ممارساتهم التقويمية فيسهل عليهم معالجتها. وبالمحصلة يستفيد أعضاء هيئة التدريس في تحسين ممارساتهم التقويمية إلى الحد الذي يؤثر بصورة إيجابية على دقة الدرجات.

٢ - تغير طبيعة المسابقات وخصائص أعضاء هيئة التدريس بين الفصلين. وتوفر الدراسات التحليلية لنتائج الطلبة في الفصلين (شهادة، وآخرون،

١٩٩٨، أ، ١٩٩٨ب) الأدلة التي تعزز الاحتمال الأول. فمن ناحية أوضحت نتائج الدراستين أن نسبة حالات التطرف في توزيع التقديرات قد انخفضت في الفصل الثاني عنها في الفصل الأول. وسواء انطوت حالات التطرف على المغالاة أو التشدد في إعطاء التقديرات المرتفعة، فإنها تؤدي إلى انخفاض مدى الدرجات وتباينها، وبالتالي ينخفض ثباتها (Masters, 1974).

ومن ناحية أخرى، فقد ازدادت قوة العلاقة بين مكونات الدرجة النهائية في الفصل الثاني عنها في الفصل الأول. ويتوقع أن ازدياد قوة العلاقة نتج بفعل التحذيرات التي نبهت إليها الدراسة التحليلية الأولى للنتائج والتي تقضي بضرورة الحد من التدخلات غير الموضوعية للمدرس في عملية تحديد أعمال السعي. وازدياد قوة العلاقة بين مكونات الدرجة النهائية يؤدي إلى زيادة قيمة الثبات حيث تقوم جميع معاملات ثبات الاتساق الداخلي على الارتباطات الداخلية لمكونات الدرجة؛ (Feldt and Brennan, 1989).

وفيما يتعلق بالأمر الثاني، فإنه لم يكن بالإمكان في هذه الدراسة ضبط المتغيرات المتعلقة بالمساق وبعضو هيئة التدريس. وبالرغم من أن التغيرات في خصائص أعضاء هيئة التدريس وطبيعة المساقات بين الفصلين ليست كبيرة (شهادة وآخرون، ١٩٩٨، أ، ١٩٩٨ب)، فإن احتمالية ارتباط التحسن في معاملات الثبات بتغير طبيعة المساقات وبخصائص أعضاء هيئة التدريس قائمة. وهذا بدوره قد يلقي بظلال من الشك على التأثير الإيجابي لتحليل نتائج الطلبة على ثبات الدرجات الجامعية ويهدد الصدق الداخلي في الدراسة إلى حد ما. وفي محاولة لتعزيز هذا الصدق، قام الباحث بحساب قيم معاملات الثبات لجميع المساقات التي تكرر تدريسها من قبل عضو هيئة التدريس نفسه في الفصل الأول من العام الدراسي ١٩٩٩/٩٨. وقد بلغ عدد هذه المساقات وبعضو هيئة التدريس. وقد أشارت قيم الثبات المحسوبة إلى أن قيم معاملات الثبات للمساقات في الفصل الأول ١٩٩٨/٩٧ في ٧٢٪ من المساقات، وقد زادت قيمة الثبات في المتوسط بمقدار ٠,٠٤٧ حيث ارتفع متوسط قيم الثبات لجميع

المساقات من ٠,٦٥٥ في الفصل الأول ١٩٩٨/٩٧ إلى ٠,٧١٢ في الفصل الأول ١٩٩٩/٩٨. وتؤكد هذه النتيجة على أن هناك أثر إيجابي لتحليل نتائج الطلبة على ثبات الدرجات في المساقات الجامعية. وهذا بدوره يعزز من احتمال أن يؤدي تحليل نتائج الطلبة إلى تغيرات إيجابية في ثبات الدرجات الجامعية.

وبصورة عامة تشير بيانات الثبات إلى أن الدرجات الجامعية في كلية العلوم الإنسانية تتمتع بدرجة مقبولة من الثبات في أكثر من نصف المساقات (٥٨٪). وتماثل هذه النتيجة ما توصل إليه سوامه (١٩٩٥) في جامعة اليرموك/ الأردن. وربما تعزى هذه النتيجة إلى اعتماد جامعة الإمارات على الاختبارات كوسيلة مهمة في تقويم عمل الطالب، إذ يصل وزنها في كثير من المساقات إلى ٧٠٪ من الدرجة الكلية. وينسجم هذا مع ما تؤكد عليه الدراسات السابقة (غوني، ١٩٩٢) من أن الاختبارات من أفضل أساليب التقويم في إصدار الحكم على التقدير النهائي للطالب. وعلى العموم يتوقع أن يكون الثبات أعلى ما يمكن في الاختبارات الموضوعية التي تبذل عناية خاصة في بنائها، ويكون الثبات متوسطاً نوعاً ما في الاختبارات المقالية والتقارير والأوراق، وأقل ما يمكن في المشاركة الصفية وأشكال التعبير الشفوي. مثل هذه الاعتبارات يجب أن تؤخذ بعين الاعتبار مستقبلاً في تحديد الأوزان المناسبة لكل أسلوب من هذه الأساليب من أجل الحصول على درجات ذات ثبات مرتفع.

وتوضح نتائج الدراسة، أن دلالة الفروق في توزيع الثبات بين الفصلين تختلف لدى البنين عنها لدى البنات، كما تختلف من قسم لآخر. فقد كانت للفروق دلالة إحصائية في شعب البنين عند مستوى الدلالة الإحصائية ($\alpha = 0,05$) بينما لم تكن لها دلالة إحصائية في شعب البنات. وكانت للفروق دلالة إحصائية في أقسام التاريخ والجغرافيا والفلسفة بينما لم يكن لها دلالة إحصائية في الأقسام الأخرى. ويرتبط التحسن في الثبات بين الفصلين بالتغيرات الإيجابية في توزيع التقديرات الجامعية. فقد أشارت نتائج الدراسات التحليلية لنتائج الطلبة (شهادة وآخرون، ١٩٩٨ أ، ١٩٩٨ ب) إلى حدوث تغيرات

واضحة لتوزيع التقديرات الجامعية بين الفصلين تمثلت في: تساهل قسم التاريخ في إعطاء التقديرات المرتفعة بعد أن كان في الفصل الأول هو القسم الأكثر تشدداً، وتشدد قسم الفلسفة في إعطاء التقديرات المرتفعة بعد أن كان هو القسم الأكثر تساهلاً. وتحلي قسم الجغرافيا عن المرتبة الأولى في نسبة التقديرات المتوسطة حيث انخفضت من المستوى المرتفع جداً (٤٦٪) إلى مستوى أكثر اعتدالاً (٤٢٪)، واقترب توزيع التقديرات لدى البنين بصورة كبيرة من التوزيع المعتدل بعد أن كان سالب الالتواء نتيجة تردد بعض أعضاء هيئة التدريس في ترسيب الطلاب. وكل هذه التغيرات تسهم في رفع تباين الدرجات وزيادة مداها، وبالتالي يتحسن ثباتها (Allen and yen, 1979). ونظراً لأن تحليل نتائج الطلبة يساعد أعضاء هيئة التدريس في تحسين ممارساتهم التقويمية إلى الحد الذي يعزز ثبات الدرجات الجامعية فإنه ينصح بالاستمرار في تحليل نتائج الطلبة في المساقات الجامعية وتعميمها على الكليات الأخرى في الجامعة.

المراجع

المراجع العربية:

- أبو لبدة، سبع (١٩٨٧). مبادئ القياس النفسي والتقييم التربوي، الطبعة الرابعة. عمان، جمعية عمال المطابع التعاونية.
- أبو علام، رجاء والصراف، قاسم (١٩٨٥ أ). دراسة تحليلية لتوزيع تقديرات الطلبة في كلية التربية للفصل الدراسي الأول ١٩٨٣/٨٢. المجلة التربوية، الكويت، العدد السابع عشر.
- أبو علام، رجاء والصراف، قاسم (١٩٨٥ ب). توزيع تقديرات الطلبة في جامعة الكويت ومضامينها الإرشادية. المجلة العربية للعلوم الإنسانية، الكويت، العدد السابع عشر.
- الطويل، هاني (١٩٨٦). الإدارة التربوية والسلوك المنظم. عمان: مطبعة كتابكم.
- المسند، شيخة (١٩٩٢). دراسة تحليلية لتوزيع تقديرات الطلاب في جامعة قطر في الفصل الأول ١٩٩٠. حولية كلية التربية، جامعة قطر، العدد التاسع، ١٢٥ - ١٥٣.
- المطوع، حسين (١٩٨٨). العلاقة بين مستوى التحصيل في الشهادة الثانوية العامة ومستوى التحصيل في الشهادة الجامعية لخريجي وخريجات الانتساب الموجه في جامعة الإمارات العربية المتحدة. حولية كلية التربية، جامعة الإمارات العربية المتحدة. العدد ٣، ١١ - ٢٩.
- بوطانة، عبدالله (١٩٨٦). دور التقويم في تطوير العملية التربوية. التربية الجديدة، مكتب اليونسكو الإقليمي للتربية في البلاد العربية، العدد ٣٩.

- ثورندايك، روبرت وهيجن، اليزابيث (١٩٨٩). القياس والتقويم في علم النفس والتربية، الطبعة الرابعة. (ترجمة عبدالله الكيلاني وعبدالرحمن عدس). مركز الكتب الأردني.
- سرحان، الدمرداش (١٩٨١). المناهج المعاصرة. الكويت، مكتبة الفلاح.
- سوامه، يوسف (١٩٩٥). تقدير الثبات للعلامات الجامعية في عينة من المساقات في جامعة اليرموك خلال الفصل الثاني ١١٩٣/٩٢. مجلة مركز البحوث التربوية، قطر، العدد السابع، ٧١ - ٨٩.
- سوامه، يوسف (٢٠٠٠). المعاني والقيم في تقديرات أعضاء هيئة التدريس للعلامات الجامعية: دراسة حالة لجامعة اليرموك. مجلة مركز البحوث التربوية، قطر، العدد ١٧، ٤٥ - ٧٧.
- شحادة، نعمان وآخرون (١٩٩٨ أ). نتائج الفصل الأول من العام الدراسي ١٩٩٨/٩٧: دراسة تحليلية. كلية العلوم الإنسانية والاجتماعية، جامعة الإمارات العربية المتحدة.
- شحادة، نعمان وآخرون (١٩٩٨ ب). نتائج الفصل الثاني من العام الدراسي ١٩٩٨/٩٧: دراسة مقارنة مع نتائج الفصل الأول. كلية العلوم الإنسانية والاجتماعية، جامعة الإمارات العربية المتحدة.
- عجايوي، محمود وخضر، فخري (١٩٨٧). دراسة تحليلية لتوزيع تقديرات الطلبة في جامعة الإمارات العربية المتحدة. حولية كلية التربية، جامعة الإمارات العربية المتحدة، ٢، ٤٥ - ٦٨.
- عودة، أحمد (١٩٩٣). القياس والتقويم في العملية التدريسية. إربد، الأردن: دار الأمل.
- غوني، منصور (١٩٩٢). كيفية إصدار حكم موضوعي على درجات الطلاب باستخدام أساليب تقويم مختلفة والتعرف على أفضلها علاقة بالتقدير النهائي للطلاب. المجلة العربية للتربية، ١٢ (٢)، ٧٠ - ٩١.
- فرحات، سهير (١٩٩٧). تقويم الطلبة لأداء أعضاء هيئة التدريس. الرياض، مكتب التربية العربي لدول الخليج.

ثانياً - المراجع الأجنبية:

- Allen, N. and Yen, W. (1979). **Introduction to Measurement Theory**. Monterey, CA: Brooks/Cole Publishing Company.
- Barnes, S. (1985). A study of classroom pupil evaluation: The missing link in teacher education. **Journal of Teacher Education**, 36 (4), 46-49.
- Bligh, D. (1988). **Higher Education**. London: Cassel Educational Limited.
- Clark, E. (1950). Reliability of college grades. Paper presented at the meeting of American Psychological Association.
- Davidson, J. (1975). Academic interest rates and grade inflation. **Education Record**. 56, 122-125.
- Duke, J. (1983). Disparities in grading practice, some resulting inequities, and a proposed new index of academic achievement. **Psychological Reports**. 53, 1023-1080.
- Elton, L. and Laurillard, D. (1979). Trends in research on student learning. **Studies in Higher Education**, 4 (1), 87-102.
- Etaugh, A., Etaugh, C. and Hurd, D. (1972). Reliability of college grades and grade point averages. **Educational and Psychological Measurement**, 32, 1045-1050.
- Feldt, L. and Brennan, R. (1989). Reliability. In R. Linn (Ed.) **Educational Measurement**. (3rd edition). New York: Macmillan Publishing Company.
- Frisbie, D. (1983). Assigning course grades. **Technical Bulletin #5**, Evaluation and examination Service, University of Iowa.
- Gronlund, N. (1985). **Measurement and Evaluation in Teaching**, (5th ed.). New York: Macmillan Publishing Company.
- Manke, M. and Loyd, B. (1991). A study of teachers understanding of their grading practices. Paper presented at the annual meeting of NCME, Chicago.
- Master, J. (1974). The relationship between number of response categories and reliability of Likert-type questionnaires. **Journal of Educational Measurement**, 11, 49-53.
- Merwin, J. (1989). Evaluation. In M. Reynolds (ed.). **Knowledge base for the beginning teacher**. Oxford: Pergamon Press.

- Noble, J. (1991). **Improving College Grade Predictions Using Expanded High School Course Work and ACT Scores**. ACT Research Report. Iowa City, IA: ACT.
- Payne, D. (1974). **The Assessment of Learning: Cognitive and Affective**. London: D.C. Health and Company.
- Raju, N. (1977). A generalization of coefficient alpha. **Psychometrika**, 42, 549-565.
- Sawyer, R. (1989). Validating the use of ACT scores for remedial course placement in college. ACT Research Report, Iowa city: ACT.
- Schoenfeldt, L. and Brush, D. (1975). Patterns of college grades across curricular areas. **American Educational Research Journal**, 12, 313-321.
- Smith, E. (1979). Combining student test scores. **Educational Research and Methods**, 11 (2), 42-50.
- Stiggins, R., Frisbie, D. and Griswold, P. (1989). Inside high school grading practices. **Educational Measurement: Issues and Practice**, 8, 6-14.
- Suslow, S. (1976). **A report on an interinstitutional survey of undergraduate scholastic grading: 1960's to 1970's**. Office of Institutional Research, UCA, Berkeley.
- Terwilliger, J. (1989). Classroom standard setting and grading practice. **Educational Measurement: Issues and Practice**, 8, 15-19.
- White, S. (1975). Sometimes an A is really an F. **Chronicle of Higher Education**, 9, p. 24.

THE EFFECT OF COURSE GRADES ANALYSIS ON COURSE TOTAL SCORES RELIABILITY

Dr. Yousef M. Sawalmeh

*Department of Psychology
United Arab Emirates University*

Abstract

The purpose of the study was to determine the effect of course grades analysis on course total scores reliability. The data for the study consisted of student's total scores and subscores in 1087 sections of courses taken during the 1997/1998 academic year at the Faculty of Humanities and Social Sciences/UAE University. Student's grades in 530 sections were assigned during the first semester before conducting the first study of course grades analysis. Student's grades in the remaining 557 sections were assigned during the second semester after conducting the study and utilizing the results of the analysis in improving evaluation practices. The reliability of course total scores in every section was assessed by applying the Raju' formula for estimating the reliability.

The findings indicated that course total scores enjoy a high level of reliability; in 58% of the sections the reliability estimate was greater than 0.70. Also, there are significant differences in reliability distributions between the two semesters; the proportion of courses enjoying high level of reliability increased due to the positive effect of course grade analysis from 53% in first semester to 62% in second semester.