

# DEVELOPPEMENT ET VALIDATION D'UNE NOUVELLE ECHELLE DE MESURE DE L'ATTITUDE A L'EGARD DES STATISTIQUES : SATS-R

Kevin Carillo<sup>1</sup> & Nadine Galy<sup>2</sup> & Cameron Guthrie<sup>3</sup> & Anne Vanhems<sup>4</sup>

<sup>1,2,3,4</sup> *Université de Toulouse, Toulouse Business School, 20, bd Lascrosses, 31068 Toulouse*

<sup>1</sup> [k.carillo@tbs-education.fr](mailto:k.carillo@tbs-education.fr),

<sup>2</sup> [n.galy@tbs-education.fr](mailto:n.galy@tbs-education.fr)

<sup>3</sup> [c.guthrie@tbs-education.fr](mailto:c.guthrie@tbs-education.fr)

<sup>4</sup> [a.vanhems@tbs-education.fr](mailto:a.vanhems@tbs-education.fr)

**Résumé.** De nombreuses études ont montré l'importance de l'attitude lors des apprentissages. Des étudiants développant une attitude positive auront tendance à obtenir de meilleurs résultats académiques que les autres. Dans le cadre de l'enseignement de la statistique, il existe plusieurs instruments permettant de mesurer cette attitude, mais le plus utilisé reste celui développé par Schau (Survey Attitude Toward Statistics, 1995, 2003). Malgré son usage répandu, la littérature montre que cet instrument reste à améliorer du point de vue de sa validité et de sa fiabilité. Dans cet article, nous revisitons l'instrument de Schau et proposons une nouvelle version plus robuste aux tests de validité et fiabilité.

**Mots-clés.** Enseignement des statistiques, Survey Attitudes Towards Statistics (SATS), Validation d'instrument de mesure

**Abstract.** Numerous studies have shown the importance of attitude during learning. Students developing a positive attitude will tend to have better academic results than others. Within the framework of statistical teaching, there are several instruments for measuring this attitude, but the most widely used is the one developed by Schau (Survey Attitude Toward Statistics, 1995, 2003). Despite its widespread use, the literature indicates that this instrument remains to be improved in terms of its validity and reliability. In this article, we revisit the Schau instrument and propose a new more robust version of the validity and reliability tests.

**Keywords.** Business statistics education, Survey Attitudes Towards Statistics (SATS), Validation of measure instrument

## 1. Introduction

Dans la littérature, l'importance de l'attitude dans le contexte d'un cours introductif aux statistiques est largement reconnue (Gal I., Ginsburg L. et al. 1997, Leong J.L. 2006). Il apparaît ainsi que l'attitude développée par les élèves va exercer un impact sur leur comportement académique, leur façon d'aborder le cours et au final sur leur réussite à l'examen (Ramirez C., Schau C. et al. 2012). Des attitudes négatives sont un obstacle majeur pour un apprentissage efficace (Waters L.K., Martelli T. A. et al. 1988).

Il existe une abondante recherche sur l'attitude des étudiants à l'égard des statistiques, mais on ne relève pas vraiment de consensus sur la façon dont elle doit être conceptualisée. Cette absence de consensus s'illustre dans les variations importantes qu'il existe dans les questionnaires conçus pour mesurer cette attitude. Ramirez et al. (2012) et Nolan et al. (2012) présentent dans une revue de littérature 15 mesures, utilisées entre 1980 et 2007 pour des recherches publiées dans des revues académiques ou présentées à des conférences. L'instrument le plus utilisé reste l'instrument développé par Schau (Survey Attitude Toward Statistics, 1995, 2003). Cependant la littérature (Nolan et al., 2012 ; Vanhoof et al., 2012) montre que cet instrument reste à améliorer du point de vue de sa

validité et de sa fiabilité.

L'objectif de notre étude est de revisiter l'instrument de Schau et de proposer une nouvelle version plus robuste aux tests de validité et fiabilité.

## 2. L'instrument SATS

### 2.1 Présentation

Schau (Schau C., Stevens J. et al. 1995), a développé le questionnaire SATS-28 (Survey of Attitudes Toward Statistics) suivi par le questionnaire SATS-36 (Schau C. 2003) qui complète les construits précédents avec deux nouvelles composantes.

Le SATS-28 contient 28 items mesurant 4 composantes de l'attitude :

- **affect**  (6 items) : sentiments envers les statistiques, plus le score est élevé, plus le sentiment est positif
- **cognitive competence**  (6 items) : perception de la compétence intellectuelle et la capacité à appliquer les statistiques, plus le score est élevé, plus la perception est forte.
- **value**  (9 items) : perception de l'utilité, de la pertinence et de la valeur des statistiques dans la vie professionnelle et personnelle, plus le score est élevé, plus la perception est forte
- **difficulty**  (7 items) : perception de la difficulté des statistiques comme sujet d'étude, plus le score est élevé, plus la perception de la difficulté est faible.

Les deux composantes supplémentaires du SATS 36 sont les suivantes :

- **interest**  (4 items) : niveau d'intérêt envers les statistiques, plus le score est élevé, plus l'intérêt est fort.
- **effort**  (4 items) : quantité de travail que l'étudiant prévoit de mettre en œuvre dans le cours de statistique, plus le score est élevé, plus l'effort prévu est important.

Chaque item se présente sous la forme d'une échelle de Likert à 7 modalités, codées de 1 (Pas du tout d'accord) à 7 (Tout à fait d'accord), 4 correspondant à Ni en accord, ni en désaccord.

Il existe deux versions du questionnaire : une version préliminaire du questionnaire (pré-SATS) doit être administrée au tout début du premier cours. Les étudiants se « projettent » dans le cours et le temps utilisé pour les différents items est le futur. La version « post » est distribuée à la fin du dernier cours, et permet de mesurer l'attitude a posteriori. Les items restent identiques, seuls les temps employés dans la formulation des questions changent d'un questionnaire à l'autre (le futur est employé dans la version pré et le passé dans la version post).

La version complète du SATS-36 est consultable en ligne : <http://evaluationandstatistics.com>.

### 2.2 Critiques de l'instrument

Plusieurs recherches se sont focalisées sur l'étude de l'instrument SATS (28 ou 36) et se sont montrées critiques relativement à sa fiabilité ou sa validité.

Les principales critiques concernent l'*item parcelling*, la validité du contenu, la validité structurelle, et la validité interne.

#### **Regroupement d'items (*item parcelling*)**

Les travaux de référence sur l'instrument SATS (Schau, Stevens et al. 1995; Dauphinee, Schau et al. 1997; Gribbons et Hocevar 1998) utilisent une technique de « *item parcelling* ». Elle consiste à regrouper pour chaque dimension certains items en groupes (*parcels*) de 3 ou 4 items en faisant leur moyenne ou leur somme. Cette stratégie a été suggérée par les auteurs dans le but de réduire les problèmes de non-normalité de certains items et d'améliorer la fiabilité. Elle reste cependant

controversée. Elle peut en effet être source de biais dans la mesure où il existe plusieurs méthodes pour effectuer les regroupements et le choix de la méthode peut affecter les résultats. De plus, elle suppose que les items regroupés en *parcels* mesurent la même dimension, ce qui n'a pas forcément été étudié ou vérifié (Bandalos et Finney, 2001). Si ce n'est pas le cas, agréger le score des items n'a aucun sens (Kline, 2005). Enfin, l'utilisation des items individuels (*unparceled*) ou regroupés (*parceled*) peut conduire à des résultats différents dans les études factorielles exploratoires ou confirmatoires.

### **Validité du contenu**

Pour le questionnaire SATS-28, Nolan et al. (2012) observent que le processus de construction et de définition des items a été très bien détaillé. Ils relèvent que parmi les principales échelles d'attitude, SATS-28 est celle qui est la plus rigoureusement développée et qui paraît représenter ainsi le plus fidèlement la population des items mesurant l'attitude envers la statistique. Cependant ils font aussi remarquer que les auteurs ne précisent pas si le même processus rigoureux a été adopté pour créer les deux mesures complémentaires (*Effort* et *Interest*) car il n'existe pas de publication disponible sur ce sujet.

### **Validité Structurelle**

Concernant la validité convergente des deux échelles SATS-28 et SATS-36, il n'y pas de consensus clair. Nolan et al. (2012) précisent que c'est vraisemblablement dû au fait que les 3 construits *Affect*, *Cognitive competence*, et *Difficulty* sont moyennement à fortement corrélés selon les études..

Une analyse factorielle confirmatoire (AFC) effectuées sur les scores de SATS-28 dans lesquels les items ont été traités en « *parcel* » a bien établi l'existence de 4 facteurs (Dauphinee et al., 1997 ; Schau et al., 1995). Mais, par ailleurs, une analyse factorielle exploratoire a révélé l'existence de seulement deux facteurs, les 3 composantes *Affect*, *Cognitive competence*, et *Difficulty* étant regroupées sur un facteur unique (Cashin et Ellmore, 2005). En ce sens l'échelle de mesure se résume à deux facteurs, l'un traduisant l'utilité accordée au cours de statistique (*Value*) et l'autre « comment un étudiant ressent le cours » (*Affect*, *Difficulty*, *Cognitive competence*). De plus, concernant SATS-36, si l'analyse confirmatoire sur les *parceled items* a vérifié la présence de six facteurs, le même type d'analyse appliqué aux données *unparceled* (items individuels) a révélé que le modèle à 4 facteurs (*Affect*, *Cognitive competence*, et *Difficulty* resortant sur une même dimension) donnait d'aussi bons résultats que le modèle à 6 facteurs (Vanhoof et al., (2012), Carillo et al. (2016)).

### **Validité interne :**

Si la validité interne mesurée par l'alpha de Cronbach est généralement bonne, elle reste parfois faible pour la composante *Difficulty* (0,51 à 0,91 selon les études) et *Value* (0,58 à 0,93 selon les études) (Cahen et al., 2005 ; Chiesi et Primi, 2009 ; Hilton et al., 2004 ; Carlson et al., 2011).

## **3. Méthodologie**

La Figure 1 résume les différentes phases et étapes suivies lors du développement et la validation de l'échelle.

### **Phase 1 : Développement et raffinement de l'échelle**

La première partie de ce projet de recherche a consisté à identifier de manière systématique les problèmes à la fois conceptuels et de validité (convergente et discriminante) de l'échelle SATS d'origine. Pour cela, nous nous sommes inspirés des conseils et de la méthode prônés par Moore et Benbasat (1991) qui est un condensé des meilleures pratiques du développement d'échelles de mesure dans les disciplines du marketing et des systèmes d'information.

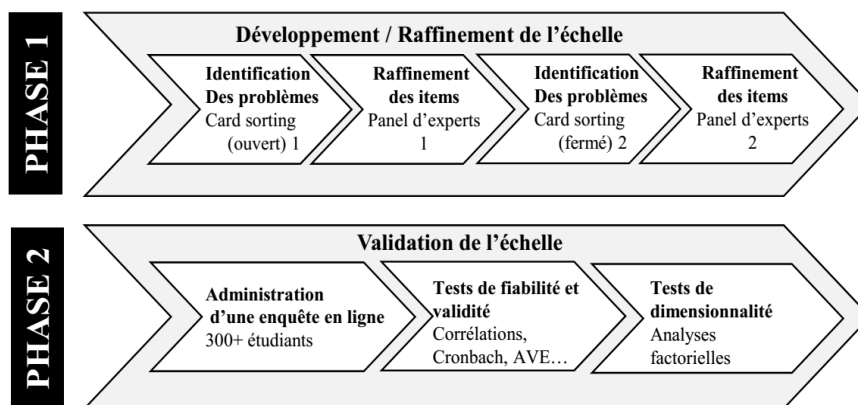


FIGURE 1 – Les différentes étapes du développement et validation de l'échelle SATS-R

#### *Première procédure de triage de cartes (card sorting) – ouverte*

Les items de mesure de l'instrument SATS original ont été imprimés chacun sur une carte. Cinq enseignants-chercheurs (dans les domaines du marketing, de la stratégie et des ressources humaines) ayant une expertise dans le développement d'échelles de mesure, ont dû regrouper les items selon les catégories de leur choix (qu'ils ont dû nommer et définir) - procédure de triage de cartes – ouverte). Les experts avaient aussi pour instructions d'indiquer les items qui paraissaient ambigus ou bien difficiles à comprendre.

#### *Panel d'experts 1 – Raffinement des items*

Un panel d'experts impliquant l'équipe de recherche mais aussi 3 professeurs et une étudiante invités se sont réunis afin de discuter des résultats. Dans un premier temps, Par exemple, les noms et définitions des catégories procurées par les cinq juges ont été comparés aux 6 dimensions initiales de SATS afin de s'assurer de la cohérence conceptuelle entre les définitions des dimensions et les items censés les mesurer. Par ailleurs, en confrontant les résultats et remarques de chacun des cinq juges, un certain nombre de problèmes ont été identifiés : des incohérences dans le sujet des items ('moi' versus 'les autres'), les temps grammaticaux utilisés, des items biaisés (non-centrés) à cause de certains termes utilisés, ou bien des allusions ambiguës à l'apprentissage et à l'utilisation des statistiques. La procédure s'est terminée par le raffinement des items posant problèmes et la génération de nouveaux items dans certains cas.

#### *Deuxième procédure de triage de cartes (card sorting) – fermée*

Les nouveaux items, imprimés eux aussi sur des cartes, ont été fournis à cinq nouveaux juges (enseignants chercheurs dans les domaines des systèmes d'information, la comptabilité, le marketing, et les ressources humaines). Lors de cette deuxième procédure, les juges ont été informés du nombre, du nom et de la définition des différentes catégories (procédure de triage de cartes – fermée). De la même manière que lors de la première procédure, les juges ont attribué une catégorie à chaque item lorsque cela leur paraissait clair, et ont mis de côté les items ambigus ou posant problème.

#### *Panel d'experts 2 – Raffinement final des items*

L'équipe de recherche s'est réunie une nouvelle fois afin de discuter des résultats de la procédure de triage de cartes 'fermée' avec pour objectif de raffiner de nouveau les items. Le détail des modifications des différents items à chaque itération du processus de développement/raffinement sera présenté dans la version longue de l'article.

### **Phase 2 : Validation de la nouvelle échelle SATS-R**

Afin d'évaluer les propriétés psychométriques de la nouvelle échelle SATS-R, une enquête en ligne

a été conduite auprès d'un public d'étudiants ayant suivi le cours d'introduction aux statistiques de première année (niveau L3) à Toulouse Business School durant l'année universitaire 2016-2017.

Les phases pré et post du questionnaire ont été administrées respectivement lors du premier et du dernier cours du module en présence d'un professeur. Le questionnaire a été administré dans sa version anglaise, les étudiants de TBS ayant un très bon niveau en anglais (le niveau de langue est un critère de sélection important lors du concours d'entrée).

La version préliminaire de SATS (version-pré) compte ainsi 312 réponses exploitables qui ont permis de valider les composantes de l'attitude.

L'échantillon compte 53,7% de filles et 46,3% de garçons. 86,5% des étudiants interrogés proviennent de classes préparatoires, les 13,5% restant ayant suivi un cursus bac+2 type DUT, BTS, L2 ...

## 4. Résultats

Une fois collectées, les données ont été analysées avec les logiciels IBM SPSS (version 19) et SmartPLS version 3.0 (Modélisation d'équations structurelles par l'approche PLS ou régression des moindres carrés partiels). Les Tableaux 3, 4, et 5 résument les principaux résultats lors de la phase de validation. La fiabilité et validité convergente des différentes dimensions ont été testées selon les tests suivants :

- Cronbach's alpha > 0,7
- Composite Reliability (CR) > 0,7
- Average Variance Extracted (AVE) > 0,5
- Item-item corrélations > 0,3 pour chacune des dimensions

La validité discriminante des dimensions a été validée grâce aux tests suivants :

- Test de Fornell -Larcker: racine carrée de l'AVE de chaque dimension > corrélations entre les dimensions
- Cross-loadings des items < loadings des items dans leur dimension respective
- Ratios HTMT (heterotrait-monotrait ratio of correlations) < 0,85

Trois items (un pour *Value* et deux pour *Difficulty*) ont été éliminés car ne ils ne satisfaisaient pas les tests mentionnés ci-dessus. L'analyse des résultats de la dimension *value* incluant originellement des items concernant la valeur des statistiques dans le cadre de la vie professionnelle mais aussi personnelle, a révélé une structure dimensionnelle non satisfaisante. Par exemple, le critère de validité convergente AVE > 0,5 n'était pas respecté, ni celui des ratios HTMT associés. Des analyses factorielles exploratoires ont permis de conclure que *value* était 'mieux' modélisée (les tests statistiques de validité convergente/discriminante le confirmant) en tant que construit formatif de second-ordre. Cela signifie que la dimension *Value* se définit en tant que l'agrégation de la perception des statistiques pour la vie professionnelle à celle de la vie personnelle.

## 5. Conclusion

Nous avons présenté ici une version revisitée de l'instrument de Schau, largement utilisé dans la littérature pour mesurer l'attitude des étudiants envers la statistique. Nos résultats ont permis de valider cette nouvelle échelle et de montrer l'amélioration des propriétés psychométriques par rapport à la version originale.

Ainsi, grâce à un décryptage des composantes essentielles qui constituent l'attitude des étudiants, nous proposons un outil permettant l'implémentation d'une pédagogie dynamique basée sur le pilotage de l'attitude. L'outil a d'abord été testé sur un cours de statistique, mais a pour objectif d'être ensuite transposable à d'autres disciplines du management.

## 6. Bibliographie

- [1] Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: New developments and techniques*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- [2] Cashin, S. E., & Elmore, P. B. (2005). The Survey of Attitudes Toward Statistics scale: A construct validity study. *Educational and Psychological Measurement*, *65*, 509-524.
- [3] Carillo, K., N. Galy, C. Guthrie et A. Vanhems (2016), « J'aime pas les stats ! » Mesure et analyse de l'attitude à l'égard du cours de statistique dans une école de Management, *Statistique et Enseignement*, *7*(1), 3-31, <http://www.statistique-et-enseignement.fr>
- [5] Chiesi, F., & Primi, C. (2009). Assessing statistics attitudes among college students: Psychometric properties of the Italian version of the Survey of Attitudes toward Statistics (SATS). *Learning and Individual Differences*, *19*, 309-313.
- [6] Dauphinee, T. L., C. Schau, C. et J.J. Stevens (1997), Survey of Attitudes Toward Statistics: Factor structure and factorial invariance for women and men, *Structural Equation Modeling: a multidisciplinary journal*, *4*(2), 129-141.
- [7] Gal, I., L. Ginsburg et C. Schau (1997), Monitoring attitudes and beliefs in statistics education, *The assessment challenge in statistics education*, 37-51.
- [8] Gribbons, B. G. et D. Hocevar (1998), Levels of aggregation in higher level confirmatory factor analysis: Application for academic self-concept, *Structural Equation Modelling*, *5*(4), 377-390.
- [9] Hilton, S.C., C. Schau et J.A. Olsen (2004), Survey of Attitudes Toward Statistics: Factor Structure Invariance by Gender and by Administration Time, *Structural Equation Modeling*, *11*(1), 92-109.
- [10] Leong, J.L. (2006), *High school students' attitudes and beliefs regarding statistics in a service- learning based statistics course*, Thèse de doctorat, Georgia State University.
- [11] Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research*, *2*(3), 192-222.
- [12] Nolan, M. M., Beran, T. & Hecker, K. G. (2012). Surveys assessing students' attitudes toward statistics: A systematic review of validity and reliability. *Statistics Education Research Journal*, *11*(2), 103-123. Retrieved from [http://iase-web.org/documents/SERJ/SERJ11\(2\)\\_Nolan.pdf](http://iase-web.org/documents/SERJ/SERJ11(2)_Nolan.pdf)
- [13] Ramirez, C., C. Schau C. et E. S. M. A. Emmioğlu (2012), The importance of attitudes in statistics education, *Statistics Education Research Journal*, *11*(2), 57-71.
- [14] Schau, C. (1992), Survey of Attitudes Toward Statistics (SATS-28), <http://evaluationandstatistics.com>.
- [15] Schau, C. (2003b), Survey of Attitudes Toward Statistics (SATS-36), <http://evaluationandstatistics.com>.
- [16] Schau, C., Stevens, J., Dauphinee, T. L. & Del Vecchio, A. (1995). The development and validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement*, *55*, 868-875.
- [17] VanHoof, S., Kuppens, S., Castro Sotos, A. E., Verschaffel, L. & Onghena, P. (2011). Measuring statistics attitudes: Structure of the Survey of Attitudes Toward Statistics (SATS-36). *Statistics Education Research Journal*, *10*(1), 35-51. Retrieved from [http://iase-web.org/documents/SERJ/SERJ10\(1\)\\_Vanhoof.pdf](http://iase-web.org/documents/SERJ/SERJ10(1)_Vanhoof.pdf).
- [18] Waters, L.K., T. A. Martelli, T. Zakrajsek et P. M. Popovich (1988), Attitudes toward statistics: an evaluation of multiples measures, *Educational and Psychological Measurement*, *48*(2), 513-516.