

**Zeitschrift:** Schweizer Erziehungs-Rundschau : Organ für das öffentliche und private Bildungswesen der Schweiz = Revue suisse d'éducation : organe de l'enseignement et de l'éducation publics et privés en Suisse

**Herausgeber:** Verband Schweizerischer Privatschulen

**Band:** 48 (1975-1976)

**Heft:** 10

**Artikel:** Analyse des objectifs et évaluation [suite]

**Autor:** Chancerel, J.L.

**DOI:** <https://doi.org/10.5169/seals-851953>

### **Nutzungsbedingungen**

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

### **Conditions d'utilisation**

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

### **Terms of use**

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

**Download PDF:** 20.02.2026

**ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>**

Institut de psychologie - Université de Neuchâtel

SEMINAIRE PEDAGOGIQUE DE MONTREUX

## Analyse des objectifs et évaluation

J. L. Chancerel

### Chapitre IV

#### DE LA DOCIMOLOGIE CLASSIQUE A L'ADAPTATION DES TESTS AUX FINALITES DE L'EVALUATION

##### Troisième partie

##### *Typologie psychométrique des tests*

##### 1. La matrice des notes

##### a) Avantage d'un modèle simple

Si nous recherchons un moyen de classer tous les tests pédagogiques, nous aurons avantage à utiliser un modèle qui se fonde sur un nombre d'hypothèses aussi réduit que possible. Nous supposerons seulement que le test est un ensemble de questions qui peuvent être cotées chacune séparément et qu'un score partiel est obtenu par l'addition de certaines de ces questions, tandis que le score total correspond à la somme de l'ensemble. On peut ramener, semble-t-il, n'importe quel moyen d'évaluation à ce schéma de base, du plus subjectif, comme la note en composition, qui ne comporterait qu'un seul item, au plus objectif où la réponse ne peut être cotée que 1 ou 0 selon quelle est juste ou fausse.

##### b) Données de base

Toutes les informations utilisables pour la mise au point ou l'interprétation des tests de connaissances peuvent être présentées sous la forme d'un tableau à double entrée où les élèves représentent les lignes et les items les colonnes. Les éléments de cette matrice représentent la note de chaque élève à chaque item. La somme des valeurs d'une ligne représente le score total d'un sujet. La moyenne d'une colonne représente la moyenne d'un item, ou le pourcentage de bonnes réponses si cet item ne peut être coté que 1 ou 0.

##### c) Le modèle additif de l'analyse de la variance

On a proposé des schémas très différents pour rendre compte de cette matrice des réponses. Nous nous en tiendrons à un modèle déterministe plutôt que stochastique et parmi les modèles déterministes, nous choisirons celui de l'analyse de la variance, encore une fois parce qu'il est plus simple que le modèle factoriel.

L'équation de base de ce modèle revient à expliquer la note donnée à un élève pour un item, comme la somme d'une moyenne générale pour l'ensemble de la matrice, plus un effet propre à la ligne (c'est-à-dire au sujet), plus un effet propre à la colonne (c'est-à-dire la question), plus un effet d'interaction entre la ligne et la colonne (facilités particulières de ce sujet pour cette question), plus, enfin, un effet dû à d'autres facteurs incontrôlés et que l'on appelle l'erreur.

On peut démontrer facilement que la somme des carrés des écarts à la moyenne générale peut être décomposée en une part due aux différences de moyenne des lignes, une deuxième part due aux différences de moyenne des colonnes, une troisième due à l'effet d'interaction et une quatrième représentant les variations dues au hasard. Nous admettrons comme connues les méthodes utilisées pour le calcul de ces sommes de carrés partielles.

Il n'est pas inutile de souligner que l'arbitraire de ce modèle réside dans l'hypothèse d'effets additifs simples. L'analyse de la somme des

carrés, par contre, résulte d'une simple identité algébrique et est indépendante de toute considération de distribution, par exemple. Dans certains cas, l'utilisation d'un test de F peut être déconseillée, mais rien n'empêche jamais de déterminer quelle est la proportion de la variance totale qui est attribuable à chaque source de variation.

##### 2. La typologie proposée

La matrice des scores, lorsque ni les sujets ni les questions ne sont structurés de façon particulière, ne permet pas le calcul séparé d'une somme de carrés pour l'erreur et d'une somme de carrés pour l'interaction. Dans le cas général, nous ne pourrions donc analyser la somme des carrés totale qu'en trois parties: la somme des carrés intersujets, la somme des carrés interquestions, et la somme des carrés interaction.

Chacune de ces sommes des carrés nous apportera une certaine information sur la part de la variabilité totale qui peut lui être attribuée. Nous aurons ainsi une information relative aux sujets, aux questions et à leur interaction.

Certaines procédures d'évaluation ne s'intéressent qu'à une seule de ces informations. D'autres totalisent deux de ces sources de variation. Les combinaisons une à une et deux à deux des trois parts de la somme des carrés totale déterminent six types de tests que nous allons examiner successivement.

Nous considérerons, en accord avec la théorie de l'information de Shan-

non, qu'une échelle de mesure est d'autant plus informante que la variance des observations est plus grande. La précision d'une échelle correspond en effet aux possibilités de différenciation qu'elle offre. Nous chercherons donc à maximiser successivement la part de la somme des carrés totale explicable par les différences entre sujets, entre questions, etc., dans l'intention d'obtenir ainsi les mesures les plus précises pour les sources de variation correspondantes. Connaissant la variance à maximiser, il nous sera alors possible d'en déduire les propriétés psychométriques souhaitables pour chacun de ces types de tests.

### 3. Tests de classement

#### a) Variance à maximiser: S

Le premier type de test que nous examinerons correspond aussi au genre le plus répandu et à celui qui a été construit en premier, historiquement. C'est celui qui cherche à différencier le plus possible les personnes. Le but à atteindre est clair: déterminer un classement des sujets aussi précis que possible pour assurer la prise de décisions dans des problèmes de sélection, de placement, de promotion scolaire, etc. Il ne s'agit pas tant d'éviter les ex-aequo que de vraiment obtenir un classement répétable et fidèle.

On obtiendra un tel instrument de mesure si l'on parvient à rendre aussi grande que possible la part de la somme de carrés intersujets dans la somme des carrés totale, si l'on s'approche du cas-type idéal présenté au tableau IV.

Tableau IV *Matrice des données pour un test de classement typique*

1	1	1	1	1	1	6
0	0	0	0	0	0	0
1	1	1	1	1	1	6
1	1	1	1	1	1	6 Total
0	0	0	0	0	0	0
0	0	0	0	0	0	0
% B.R.	50	50	50	50	50	50 Moyenne générale

Dans une situation comme celle-ci, toute la variation entre les réponses est due aux différences entre sujets; on a annulé les variations entre questions et il n'y a aucune in-

teraction sujet-question pour troubler le classement des sujets, qui est parfaitement reproduit lorsque l'on passe d'une question à une autre.

#### b) Caractéristiques psychométriques souhaitées

On voit que la variance intersujets sera maximale si la distribution des scores totaux est binodale. Les pourcentages de réponses correctes aux questions doivent au contraire être tous identiques et égaux à 50 pour cent, car c'est cette valeur qui maximise les possibilités de variation entre sujets (aucune variation n'est plus possible à 0 % ou 100 % de bonnes réponses). Les intercorrélations entre les questions doivent tendre vers 1.00 car on souhaite que chaque item mesure la même chose que les autres et classe les sujets de la même façon.

#### c) Procédures de contrôle

Avant d'utiliser de tels tests, on contrôlera d'abord leur validité globale. Le test doit mettre en évidence des différences statistiquement significatives entre des groupes de sujets dont on sait qu'ils se distinguent sur le trait considéré. La corrélation du test avec un critère extérieur est un cas particulier de cette discrimination de groupes, correspondant au cas où les groupes sont constitués sur la base de la variable continue prise comme critère. Lorsqu'on ne possède pas de critère extérieur, on peut tester si la variance intersujets est significativement plus grande que la variance interaction.

Si le test dans son ensemble est valide, il vaut la peine de l'améliorer en sélectionnant les items les meilleurs. On conservera les questions qui discriminent le mieux entre les groupes de sujets considérés. (Ce peut être les élèves situés au-dessus ou au-dessous de la moyenne générale, si l'on n'a pas d'autre critère que la note totale du test.)

### 4. Tests d'enquête

#### a) Variance à maximiser:

Une situation différente, mais présentant un parallélisme marqué avec la précédente, est obtenue lorsque le problème n'est plus de différencier les sujets, mais les questions.

Le cas se présente dans les situations d'enquête, lorsque des chercheurs, des directeurs d'école, des autorités scolaires désirent étudier de près le fonctionnement du système scolaire. Que savent les élèves sur tel sujet? Comment se distinguent ceux qui ont suivi telle méthode d'enseignement? Les élèves de telle section atteignent-ils les objectifs fixés par leur curriculum?

Il est intéressant de remarquer que le problème est symétrique du précédent. On cherche à mettre en évidence des différences de réussite stables et répétables, mais cette fois-ci entre les questions. Les conclusions seront d'autant plus claires et nettes que ces différences seront plus grandes et qu'aucune différence entre les sujets n'obligera à nuancer ou atténuer ces conclusions relatives à la réussite aux questions. Ceci nécessite aussi que l'interaction sujets-questions soit minimum. Un cas-type idéal est présenté au tableau V.

Tableau V *Nature des données pour un test d'enquête typique*

1	0	1	1	0	0	3
1	0	1	1	0	0	3
1	0	1	1	0	0	3 Total
1	0	1	1	0	0	3
1	0	1	1	0	0	3
1	0	1	1	0	0	3
% B.R.	100	0	100	100	0	0 50 Moyenne générale

#### b) Caractéristiques psychométriques souhaitées

La distribution des scores des sujets doit présenter le moins possible de variation. Il est même souhaitable que cette tendance centrale se situe à la moitié du maximum possible, pour laisser le plus de marge de variation possible aux questions.

La distribution des pourcentages de bonnes réponses doit au contraire être bimodale, pour maximiser la variance interquestions. Les intercorrélations entre questions doivent tendre vers zéro, puisque la variation entre les sujets disparaît. Des corrélations nulles sont aussi souhaitables pour que chaque question apporte une information nouvelle. La redondance est en effet inutile dans une situation d'enquête où c'est le nombre de sujets qui réduit les

fluctuations d'échantillonnage et non le nombre de questions.

### c) Procédures de contrôle

On ne pourra tirer de conclusions assurées de l'enquête que si les différences entre questions sont statistiquement significatives. La première démarche de contrôle sera donc de tester s'il existe une différence marquée entre groupes de questions possédant des caractéristiques différentes. (On pourrait aussi tester par analyse de la variance si les différences entre questions, prises isolément, occasionnent une variance supérieure à la variance interaction. Une telle conclusion justifierait un classement des questions par ordre de difficulté, mais n'apporterait pas autant d'éléments d'interprétation que la comparaison entre groupes de questions, qui permet de tester des hypothèses précises). L'emploi de plans expérimentaux factoriels, complets ou incomplets, pour guider la rédaction des questions, facilitera beaucoup la validation globale des tests d'enquête, en suggérant les hypothèses à tester. Lorsque des facteurs apparaîtront comme significatifs, on en conclura que les paliers d'acquisitions doivent être distingués et que les objectifs pédagogiques et les tests de connaissances doivent distinguer ces cas. Les épreuves pédagogiques devront alors donner plusieurs informations distinctes, correspondant à chacun des objectifs que l'on aura distingués. Il est important de noter cet aspect multivarié fréquent des tests d'enquête.

Après cette validation globale, un certain modèle aura été construit pour expliquer les différences de réussite entre questions. Ce modèle sera constitué par les facteurs apparus comme significatifs. On pourra contrôler alors l'accord de chaque question avec le modèle, en regardant si son pourcentage de bonnes réponses est significativement différent de ce que le modèle laisserait prévoir. Un remplacement des questions déviantes sera alors possible, de même qu'une analyse psychopédagogique des raisons probables de ces déviations.

On voit que la méthodologie de la mise au point des tests d'enquête retrouve la méthode scientifique habituelle, avec ses cycles successifs de formulation et de test d'hypothèses. Si la finalité du test est réellement celle d'une enquête, l'emploi des techniques habituelles d'analyse d'items, destinées à déceler les items non-discriminatifs ne se justifie nullement et peut même conduire à une dangereuse restriction de l'information apportée par la recherche.

### 5. Tests d'inclination

#### a) Variance à maximiser:

La possibilité de maximiser la variance interaction apparaît comme une suite logique, après celle de la variance intersujets ou interquestions. Cette maximisation implique que les questions soient toutes de difficulté moyenne et que les sujets obtiennent tous des scores également éloignés des extrêmes, pour que la marge de variation la plus grande possible soit réservée aux différences de réussite à l'intérieur de chaque sujet et de chaque question.

Comme on le voit par l'exemple du tableau VI, il n'y a interaction que si les questions faciles deviennent difficiles et réciproquement lorsqu'on passe d'un sujet à un autre. — Si cette situation a peu de chances de se produire avec des élèves qui ont suivi le même curriculum, elle est normale si l'on s'adresse à des groupes d'élèves qui ont pu choisir leurs options. Les domaines étudiés et non étudiés donneront naissance à des polarités de compétence qu'il pourrait être utile de mettre en évidence. — Plus fréquemment encore, des questionnaires d'intérêt donneront lieu à une diversification des réponses du type de celle du tableau VI, parce qu'une certaine polarisation tend à se manifester, dans les motivations, opposant l'intérêt pour les choses à l'intérêt pour les personnes, par exemple. C'est pourquoi nous avons désigné ce type d'épreuves comme des tests d'inclination, pour rappeler que le niveau général de l'élève n'intervient plus dans le résultat, mais seulement la direction de ses facilités particulières.

Tableau VI *Matrice des données pour un test d'inclination typique*

1	1	1	0	0	0	3	
0	0	0	1	1	1	3	
1	1	1	0	0	0	3	Total
1	1	1	0	0	0	3	
0	0	0	1	1	1	3	
0	0	0	1	1	1	3	
% B.R.	50	50	50	50	50	50	Moyenne générale

L'importance pédagogique de ces épreuves n'est pas à sous-estimer au moment où l'on cherche à mettre chaque élève dans les conditions d'apprentissage appropriées à ses besoins propres. L'existence de styles cognitifs différents offre une voie d'approche à cette forme d'individualisation des méthodes d'enseignement. Les polarités entre visuel et auditif, verbal et pratique, intuitif et déductif, etc., devront être mesurées et exploitées, en particulier par l'enseignement assisté par ordinateur.

#### b) Caractéristiques psychométriques souhaitées

La distribution des scores totaux de sujets doit être aussi réduite que possible, nous l'avons vu, de façon que des différences de niveau général ne viennent pas troubler la manifestation de la polarité qui nous intéresse. Il en est de même de la distribution des pourcentages de bonnes réponses et les tendances centrales doivent être, dans les deux cas, également éloignées des résultats extrêmes. Les corrélations entre questions seront très élevées à l'intérieur du groupe de questions correspondant à chaque pôle, et très négatives entre questions de pôles différents.

Il est important de noter que si l'on reflète les questions de l'un des pôles, en cotant -1 la réussite à ces items, le total de la ligne devient alors l'expression de l'inclination pour l'autre pôle. Le test est ainsi transformé en un test de classement et les caractéristiques psychométriques souhaitables s'en déduisent directement.

#### c) Procédures de contrôle

La remarque précédente nous permet de passer vite sur les techniques de mise au point de ces épreuves qui

retrouvent celles des tests de classement, c'est-à-dire les méthodes d'analyse d'items habituelles. Notons, par contre, que les tests d'inclination, comme les tests de classement, sont nécessairement unidimension-

nels. Vouloir tirer de la matrice de données deux ou plusieurs scores pour chaque sujet obligerait à diminuer la variance vraie, et donc la fidélité de chaque information.

(à suivre)

## Die kreative Persönlichkeit – ihre Fähigkeiten, Eigenschaften und Motive

Prof. Dr. K. H. Delhees, Hochschule St.Gallen

(Schluß)

### b) Kreativität und Urteilsfähigkeit

Der kreative Mensch ist in großem Maße von seiner Urteilsfähigkeit hinsichtlich seiner eigenen Arbeit und ihrer Produkte überzeugt. Er hat eigene Wertmaßstäbe. Wenn er glaubt, daß seine Arbeit gut ist, kümmert er sich nicht um die Kritik anderer Personen oder setzt sich darüber hinweg bzw. wertet ihr Urteil ab. Der kreative Mensch ist nicht auf das Urteil anderer angewiesen. Er findet nämlich Ideen wichtiger als Menschen, ohne dabei etwa menschenfeindlich zu sein.

Was über die Urteilsfähigkeit kreativer Personen gesagt wurde, trägt allerdings nicht unbedingt zu ihrer Beliebtheit bei andern (auch in der Schule) bei. Sie sind deshalb in Gefahr sich von ihrer Umwelt (Eltern, Geschwister, Kameraden, Lehrer, Kollegen) zu entfremden, wozu oft das fehlende gegenseitige Verständnis beisteuert.

### c) Kreativität und Konformität

Nach Luria (1961) bewegt sich die kindliche Entwicklung um die Kontrolle von Impulsen. Auch Freud (1940) sieht darin, wie jemand seinen Impulsen nachgibt, eine wichtige Eigenart seiner Persönlichkeit. Das Kind lernt im Laufe seiner Entwicklung «Stop-Regeln» (Miller et al. 1960), die sein Verhalten in Bahnen lenken, die sich danach richten, was von den Erwachsenen in seiner Umwelt (anfänglich vor allem von den Eltern) als «richtig» erachtet wird. Sein Verhalten wird konform. Was als «richtig» oder «falsch» gilt, hängt weitgehend von der zugehörigen Kultur ab. Ein Kind, das ge-

lernt hat, in höchstem Maße beherrscht zu sein, auch im Denken, wird ausgesprochen stereotypes Denken zeigen und selten, wenn überhaupt, kreativ denken. Kreativität wird erst ermöglicht durch ein nicht zu starres Festhalten an kulturellen Regeln und Normen und der Bereitschaft, seinen Impulsen freien Lauf zu lassen, auch wenn es «gehörig» ist, sie zu beherrschen oder zurückzuhalten. In diesem Sinn lehnen hochkreative Menschen die Unterdrückung oder Verdrängung als Mechanismus der Beherrschung von Impulsen ab, was in empirischen Untersuchungen nachgewiesen ist. Mackinnon (1962) und Garwood (1964) sind zu Ergebnissen gekommen, wonach hochkreative Menschen in geringerem Maße ihre Impulse zurückhalten als nichtkreative Personen. Hudson (1966) ist der Ansicht, daß auch Personen, die weitgehend konvergent denken, zu divergentem (d. h. kreativem) Denken gebracht werden können, wenn es gelingt, sie aus ihrem stark beherrschten Verhalten (aus ihrer Konformität) zu lösen, sie also dazu zu bringen, sich «gehen zu lassen». Solche Versuche sind in gruppendynamischen Übungen von Torrance (1972) unternommen worden und zeigen (unter anderem) nach zehn Wochen Training (zwei Vormittage pro Woche) eine signifikante Verringerung der Orientierung an Vorschriften und der Abhängigkeit von Regeln.

Kreativität darf aber nicht verwechselt werden mit freiem, ungezügelter oder gar gesetzwidrigem Verhalten. Auch wenn divergieren-

des Verhalten oft aus einem Impuls heraus entsteht, so ist ein Ausleben von Impulsen noch nicht Kreativität. Auch anders sein oder etwas anders tun ist nicht in sich schon kreativ. Sicherlich kann sich ein kreativer Mensch unkonventionell verhalten oder außerhalb bestehender Normen leben, doch macht nicht die Unkonventionalität ihn kreativ; sie ist höchstens eine Begleiterscheinung, nicht aber ein integrierender Bestandteil der Kreativität.

### d) Kreativität und Risikobereitschaft

Welche und wieviel Information aus der Umwelt aufgenommen wird, ist individuell verschieden. Der eine beschränkt sich auf einen kleinen Teil, der andere holt dagegen weit aus. Die für einen Menschen charakteristische Art, Informationen aus seiner Umwelt aufzunehmen, wird als sein «kognitiver Stil» bezeichnet. Wer die maximal verfügbare Information aus seiner Umwelt aufnimmt, zeigt den kognitiven Stil der Risikobereitschaft. Der divergente Denker ist offen gegenüber der großen Vielfalt von Umweltinformation; der konvergente Denker hat im voraus eine ziemlich klare Vorstellung, was zusammenpaßt. Er weiß, was logisch ist und was nicht. In seiner Welt ist alles wohlgeordnet und am richtigen Platz.

Sich gegen alle Umweltinformationen zu öffnen, schließt aber das Risiko ein, Fehler zu machen und nicht ernstgenommen zu werden. Indem er sich nicht auf bewährte und vertraute Prinzipien und Regeln (auch der Logik) stützt, muß sich der divergente Denker ständig neuen Verhältnissen anpassen. Dabei wird er sich nicht immer – geistig wie handelnd – im Gleichschritt mit seinen Mitmenschen bewegen können, was ihm Zensur und Kritik einbringen kann. McClelland (1963) und Roe (1963) betrachten Risikobereitschaft als ein kritisches Merkmal des hochkreativen Menschen.

Eng verbunden mit der Bereitschaft, mit Ideen etwas zu riskieren, ist eine weitere Eigenschaft des kreativen Menschen: Er wagt sich an