

Le statut épistémologique de l'évaluation en éducation

Paradigmes en jeu, complexité de l'objet et défis du praticien

Eddoubi Abdelhak, Ph.D.

eddoubi@gmail.com

Plan

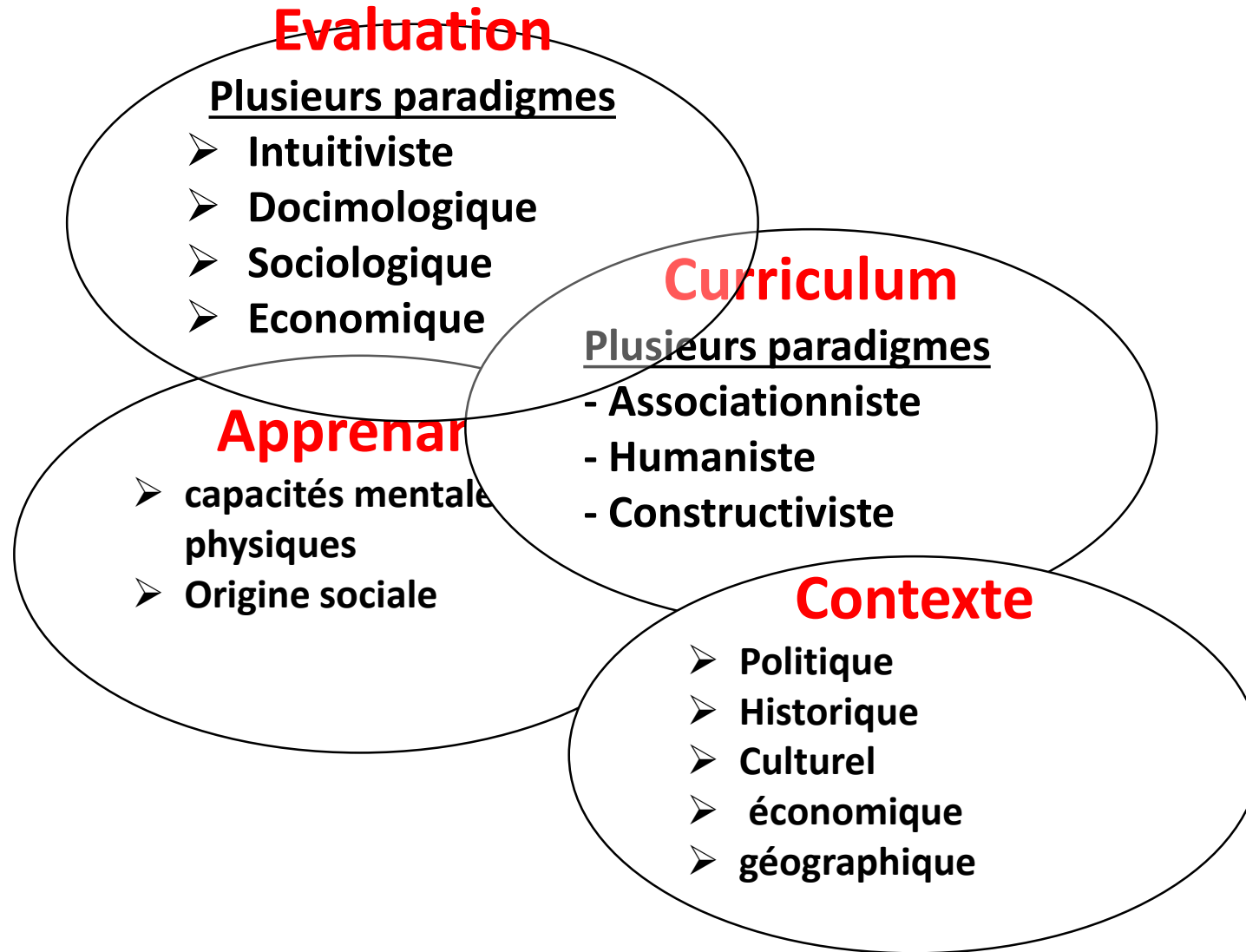
- **Poser** des questions sur la connaissance produite par l'évaluation en rappelant les apports des disciplines et la construction toujours inachevée des paradigmes qui sont à la recherche du sens de l'évaluation.
- **Situer** les conséquences prévues et inattendues qui sont sources de tensions entre les acteurs
- **Illustrer** certaines de ces tensions et relever certaines apories en s'appuyant sur l'exemple des tests standardisés (à grand enjeux et à faibles enjeux)
- **Faire quelques suggestions** pour le praticien de l'éducation

Les postulats du positivisme et du constructivisme empruntés à Lincoln & Guba (1985, p. 37)

Axiome sur	Paradigme positiviste	Paradigme constructiviste (naturaliste)
La nature de la réalité.	La réalité est unique, tangible et fragmentable.	Les réalités sont multiples, construites et holistiques.
La relation entre le connaissant et le connu.	Le connaissant et le connu sont indépendants ; dualité.	Le connaissant et le connu sont interactifs, inséparables.
La possibilité de généralisation.	Les généralisations au-delà du temps et du contexte sont possibles.	Seules les hypothèses circonscrites dans le temps et le contexte sont possibles.
La possibilité de faire les liens causals.	Il y a des causes réelles, précédant temporellement les effets ou se présentant simultanément.	Toutes les entités sont dans un état de formation simultanée mutuelle. Ainsi, il est impossible de distinguer causes et effets.
Le rôle des valeurs.	L'investigation est neutre (<i>value-free</i>).	L'investigation est liée aux valeurs (<i>value-bound</i>).

Ce sont ces deux visions de penser le monde qui expliquent le conflit paradigmatique en évaluation.

Complexité de l'évaluation



Le cadre conceptuel de la validité de Messick (1989)

1. Les éléments de preuve basés sur **le contenu d'un test** (*the content aspect*) : l'on s'intéresse à la pertinence du contenu d'un test par rapport au construit mesuré, à sa représentativité face au domaine du construit mesuré et aux qualités techniques dudit test (format des items, directives pour l'administration, phraséologie, etc.),
2. Les éléments de preuve basés sur **les processus de réponse** (*the substantive aspect*) : l'on s'intéresse à la correspondance existant entre les processus cognitifs sous-jacents au construit mesuré et ceux effectivement déployés par les répondants lors de la passation d'un test,

Le cadre conceptuel Unifié de la validité de Messick

3. Les éléments de preuve basés sur **la structure interne** (*the structural aspect*) : l'on s'intéresse à savoir si la relation entre les items et les composantes d'un test est conforme aux manifestations habituelles, connues ou souhaitées du construit qui est mesuré,
4. Les éléments de preuve basés sur **les relations avec d'autres variables** (*the external aspect*) : l'on s'intéresse à la relation entre les scores d'un test et des variables extérieures à ce test,
5. Les éléments de preuve basés sur **la généralisation de la validité** (*the generalizability aspect*) : l'on s'intéresse à savoir si les interprétations des scores d'un test peuvent être généralisées à d'autres contextes,

Les sources de tensions entre acteurs

Souscription à un paradigme qui entre en contradiction avec les résultats de l'évaluation (désaccord sur les intentions).

Application des résultats pour des fins autres que les objectifs assignés à l'évaluation

La tâche de l'évaluateur serait donc d'anticiper les conséquences de l'évaluation C'est la validité de conséquences *Consequential Validity* définie par Messick, 1995) qui consiste en ce qui suit:

“Evidence and rationales for evaluating the intended and unintended consequences of score interpretation and use in both the short- and long-term.”

L'exemple des tests à faibles enjeux

Quelques exemples: NAEP aux Etats-Unis et PNEA au Maroc

- Visées formatives
- Permet d'orienter les politiques éducatives
- Tests destinés à tester des niveaux donnés du parcours scolaire
- **Les domaines de compétences** peuvent varier selon le choix du donneur d'ordre
- Approche: évaluation à grande échelle
- Paradigme dominant: Psychométrie (*Rash* pour le NAEP et la *TCT* pour le PNEA)
- Faiblesses: **problèmes d'inférence**, interprétation, **le choix des items par des experts!!!!**
- Problème de motivation des élèves à prendre au sérieux ce genre de tests

L'exemple des tests à faibles enjeux

Quelques exemples: PISA-TIMSS-PIRLS

- **Au début, le monde est un grand laboratoire d'expérience pour élaborer une loi universelle qui explique les systèmes éducatifs**
- **Visées comparatives des systèmes éducatifs**
- **Enseignements à tirer: Pas grand-chose!**
- **Trop de biais pour tirer des leçons: culturel, dominance anglo-saxonne, format des questions**
- **Échantillonnage: beaucoup d'exclusions, sélection, inclusion**
- **Régions éloignées, élèves avec des déficiences, petites écoles!**
- **Problème de motivation des élèves pour se donner à fond**
- **Pas d'impacts proprement dit sur les enseignants et les élèves**

L'exemple des tests à grands enjeux

Quelques exemples:

- L'idée est de sélectionner et classer les élèves en les plaçant sous les mêmes conditions: même contenu, même format de tests, même temps, même correction, etc. Le souci= équité
- Quelques effets pervers:
- Rétrécissement des curricula : on concentre les programmes sur l'essentiel
- *Teaching to the Test*
- Favorise la triche
- Occulte le processus d'apprentissage
- Impacts sur la santé des élèves (stress, fatigue, dépression, suicide) et sur leur choix de carrière.
- Encourage la mémorisation

L'exemple des tests à grands enjeux

Quelques exemples de tests standardisés à grands enjeux américains:

- **Le TVAAS: Tennessee Value Added Assessment System**
- **Le Texas Assessment of Knowledge and Skills (TAKS) remplacé récemment par le State of Texas Assessments of Academic Readiness (STAAR)**

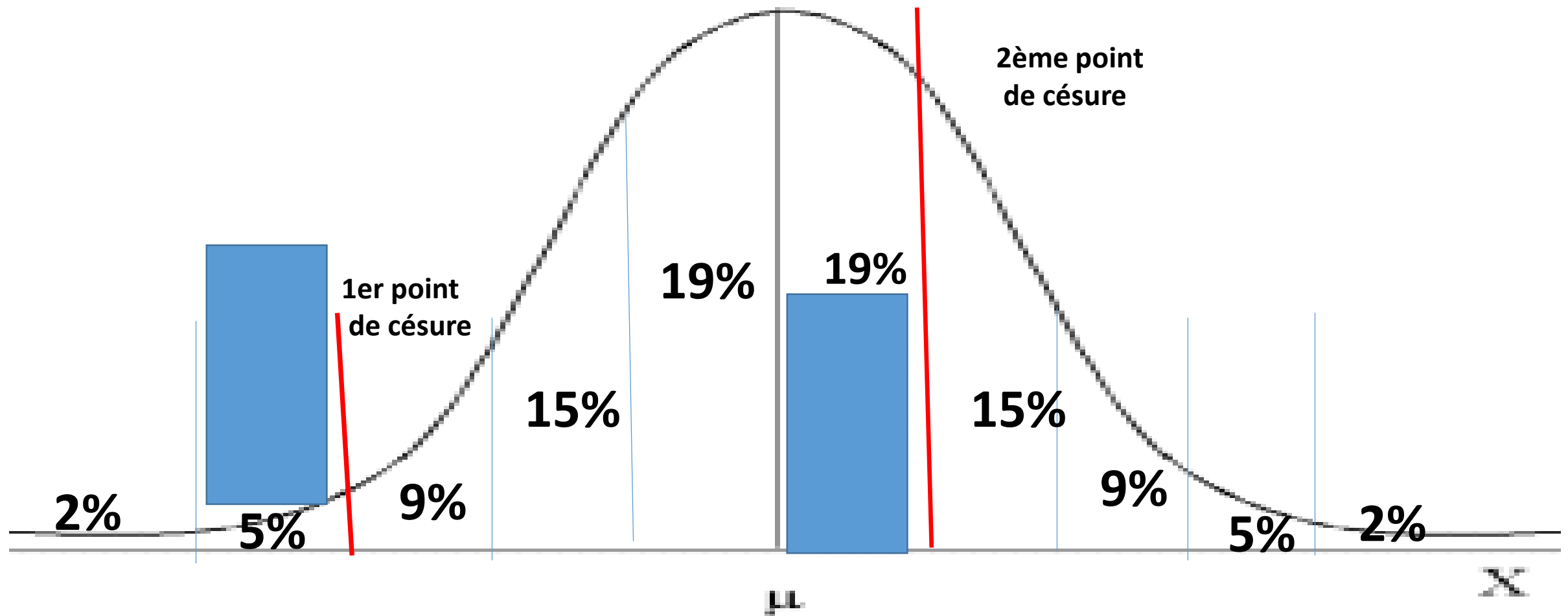
Contexte: Nation at Risk (1983) et No Child Left Behind (2001)

Modèle d'imputabilité ou de responsabilisation des établissements scolaires

New Public Management: comme mode de gouvernance du système éducatif.

Modèles de suivi de la progression des élèves utilisant les modèles hiérarchiques (Multilevels) et les Longitudinal Multilevel Analysis

Pourcentage des élèves maîtrisant les compétences sous NAEP



The Problem with Proficiency: Limitations of Statistics and Policy Under No Child Left Behind
Andrew Dean Ho, Educational researcher, Vol. 37, No 6, pp. 351-360, 2008

L'exemple des tests à grands enjeux: dérives morales ou commerce lucratifs?

1- Policy Borrowing vs Reality Learning

- Débats sur la langue d'enseignement (Afrique du Sud, Maroc)

2. Les effets pervers d'un dispositif d'imputabilité et de responsabilisation des enseignants et des établissements:

- Les scandales de la triche aux Etats-Unis, Chine, Algérie, Maroc

- *No Bush Left Behind*

L'exemple des tests à grands enjeux: dérives morales ou commerce lucratifs?

Les tests standardisés à grands enjeux est un secteur qui génère des milliards de dollars: un commerce très lucratif!!!

Rien qu'aux Etats-Unis quatre grandes entreprises accaparent la part du lion du marché du testing:

- **Harcourt Educational Measurement (HEM)** basé à Saint-Antonio au Texas et propriété du groupe d'édition britannique: Red Elsvier. HEM possède les maisons d'édition Holt, Reinhart et Winston qui sont des leaders dans l'édition des manuels scolaires du préscolaire au secondaire (K12)
- HEM est connue pour le Stanford Achievement Test (SAT-9): 15 millions d'élèves passent chaque année ce test!!!!

L'exemple des tests à grands enjeux: dérives morales ou commerce lucratifs?

CTB Mc Graw-Hill qui détient 40% du marché du testing aux Etats-Unis

Test commercialisé: Terra Nova (CTBS) est utilisé dans 19 Etats américains pour évaluer la progression des acquis scolaires des élèves.

Mc Graw-Hill qui est propriétaire de Standard & Poors, Business Week Magazine et 4 chaînes de TV avait acheté CTB en 1965. En 2000 Mc Graw-Hill avait réalisé un chiffre d'affaires de 4.2 milliards de dollars dont 34% provenait de sa division éducative.

L'exemple des tests à grands enjeux: dérives morales ou commerce lucratifs?

Reverside Publishing

Test développé: Le *Iowa Test of Basic Skills*

Part du marché: 20%

Filiale de la compagnie *Houghton Mifflin* qui a été achetée par le groupe français Vivendi Universal.

Le ITBS est administré à environ 4 à 5 millions d'élèves dans 8 états américains.

L'exemple des tests à grands enjeux: dérives morales ou commerce lucratifs?

NCS Pearson est une boîte spécialisée dans la correction des tests.

Pearson Education, une filiale du groupe britannique Pearson, basé à Londres qui est propriétaire de The Economist, The Financial Times, le Penguin Book Company, et The Network Education Family.

Chaque année, NCS corrige de près de 40 millions de tests, plus que toute autre société dans les Etats-Unis et fournit ses services à 15 états, y compris certains des plus grands marchés (Texas, la Floride, et New York).

NCS basée au Minnesota a été fondée en 1962 (il était alors connu sous le nom National Computer System). La société a été acquise par Pearson en Septembre 2000 pour 2,5 milliards de \$ et a été intégrée à la Pearson Education Unit.

En 2000 Pearson a réalisé un chiffre d'affaires de 629,5 millions \$ dont 202,4 millions \$ (32%) provenaient de sa division éducative.

Loi de Campbell

Plus on utilise un indicateur social quantitatif à des fins de prise de décision, plus il est sujet à des pressions corruptives et plus il est capable de distordre et de corrompre le processus social qu'il est censé piloter »

(Donald T. Campbell, "Assessing the Impact of Planned Social Change", Evaluation and Program Planning, II-1, 1979, p 67-90)

Implications pour le praticien

- **Bâtir les ponts entre la recherche en sciences de l'éducation et la pratique (**beaucoup de retard**)**
- **Penser à la validité des conséquences prévues et imprévues**
- **Il y n'y a pas un paradigme supérieur à un autre**
- **Décloisonner les perspectives et puiser dans la multidisciplinarité**
- **Penser à l'utilité des résultats et aux impacts sur les enseignants et les apprenants**