

## **Analyse de l'évaluation et des résultats**

### 1) Analyse des résultats des élèves

En préambule, il convient de préciser que le test ici analysé est le fruit d'une collaboration avec mon praticien-formateur, aussi bien au niveau de la conception que la correction. En effet, étant en stage A, il m'a été difficile de changer du tout au tout les habitudes de la classe à être évaluée ; c'est pourquoi une bonne partie du test consiste en des questions plus ou moins complexes. Il n'en reste pas moins que j'ai tenté de varier au maximum les approches en faisant appel aussi bien aux connaissances pures des élèves qu'aux compétences et savoir-faire entraînés en classe lors des cours.

Cela dit, dans l'ensemble, les résultats sont bons. Seuls quatre élèves sur dix-neuf n'ont pas atteint la moyenne (un 3.5 et trois 3). A part ces derniers, les notes varient entre 5.5 et 4, avec une majorité de 4.5. Ainsi la moyenne de classe résulte assez élevée : 4.39 (4.5).

En ce qui concerne le tableau de présentation des résultats ci-joint, il convient de remarquer que le seuil de réussite pour chaque exercice a été choisi a posteriori. En effet, partant de l'échelle selon laquelle le minimum de points à atteindre pour obtenir un 4 était de 27, donc environ 64% de réussite, j'ai tenté de faire coïncider ce taux avec le nombre de points à disposition pour chaque question. D'où un côté un peu arbitraire pour atteindre les 27 points requis : par exemple, le seuil de réussite de la partie B-question 1 valant 2 points a dû être abaissé à 1 point afin d'arriver à un total de 27 points, alors que toutes les autres questions rapportant 2 points ont un seuil de réussite de 1.5 points. Il s'agit probablement d'une des limites du système d'évaluation choisi pour ce test (nombre de points sans pondération = note), mais, comme mentionné précédemment, il ne m'était pas possible de modifier les habitudes d'évaluation de la classe sur une période aussi courte. J'ai donc dû « m'arranger » au mieux pour donner une cohérence au tableau d'analyse.

Partant de ce postulat l'on peut faire plusieurs observations à partir du tableau des résultats. Tout d'abord (et ceci malgré le côté arbitraire...), le nombre d'exercices sous le seuil de réussite par élève correspond généralement à la note finale (5.5 = 1-2 questions sous le seuil ; 3 = 7 questions sous le seuil). Néanmoins il convient de remarquer que certains élèves qui sont sous le seuil de réussite dans 5 questions obtiennent des résultats très divergents, allant du 3 au 5. Le problème, et il s'agit ici de la principale limite de ce test, vient du fait que les questions valant 10 points pèsent très lourd sur le résultat final : par exemple sur les trois élèves (3, 12 et 19 dans le tableau) ayant obtenu un résultat sous le seuil de réussite dans la partie A, deux d'entre eux (3 et 19) obtiennent une note insuffisante (3).

Partant de cette constatation, prenons les exemples des élèves 1, 2 et 3.

L'élève 1 n'a pas obtenu le seuil de réussite pour seulement deux questions (partie D,1 : 0/2 et partie E,2 : 1/2) ; de plus, il obtient quasiment le maximum de points pour toutes les autres

questions. Paradoxalement, la note finale ne dépasse pas 4.5. Le problème vient du fait qu'il obtient la limite du seuil de réussite pour la partie A (6/10). Ceci venant appuyer la constatation que si un des deux exercices valant 10 points n'était pas « assuré » la note finale en ressent grandement. Pour continuer dans cette réflexion, les résultats des élèves 2 et 3 sont à mettre parallèle. Alors que tous les deux n'ont pas atteint le seuil de réussite dans 5 questions, l'un obtient une note finale de 5, l'autre 3. La différence entre les deux découle encore une fois des exercices valant 10 points : l'élève 2 a obtenu 10 pts pour la partie A, alors que l'élève 3 seulement 4 pts, de même pour la partie C-question 2, l'élève 2 a obtenu 9 pts et l'élève 3, 7 pts. La différence entre ces derniers pour ces deux exercices s'élève déjà à 8 points ! Remarquant que leur total de points diffère de 10 points exactement, force est de constater que la pondération de ces deux exercices (surtout de la partie A, pure restitution de connaissances) pose problème. Nous y reviendrons dans la partie concernant l'analyse rétrospective du dispositif d'évaluation. Remarquons en outre que ces trois élèves n'ont pas atteint le seuil de réussite pour la dernière question (partie E-2).

Si l'on prend les résultats pour chaque question d'un point de vue global, mis à part le problème de pondération évoqué ci-dessus, il est intéressant de remarquer que la partie A ne fait pas partie des questions qui ont le plus posé problème (3 élèves sous le seuil de réussite seulement). En ce qui concerne la partie B, l'exercice 3 est sujet à interrogation (9 élèves sous le seuil de réussite). Il est intéressant de remarquer que l'exercice 1 de la partie C (10 élèves sous le seuil de réussite), bien que simple question de connaissance, n'a pas été une réussite ; de même la partie D, dans son ensemble, et surtout la première tâche (12 élèves sous le seuil), a posé problème. Cette partie du test était probablement la plus complexe au niveau taxonomique (analyse et synthèse de documents), bien qu'entraînée longuement en classe : cet état de fait vient corroborer mes premières réflexions selon lesquelles il est difficile, en tant que stagiaire A, de changer les habitudes d'évaluation (la classe en question ayant toujours été habituée à des tests « question-réponse »)... Finalement, nous pouvons remarquer un taux de non-réussite élevé pour la question 2 de la partie E (13 élèves sous le seuil). Les quatre questions ici évoquées seront reprises dans l'analyse rétrospective du dispositif d'évaluation.

Passons aux aspects positifs. Force est de constater, j'en veux pour preuve la bonne moyenne de classe et le faible nombre de notes insuffisantes, que de nombreuses questions ont été bien comprises et les réflexions des élèves étaient pertinentes. La partie B du test consistant en une compréhension de texte et une restitution de connaissances à partir d'un document inconnu est réussi (excepté, en partie, la question 3) : aucun élève sous le seuil de réussite pour la question 1, peu d'insuffisance pour les questions 3 et 4. Les résultats de la partie C sont paradoxaux. En effet, alors que la première question (moins bien réussie) ne faisait appel qu'à une restitution de connaissances, la question 2, plus complexe (analyse)

est presque parfaitement réussie. Bien que critiquée précédemment, la partie D du test présente de nombreuses réflexions pertinentes. S'agissant peut-être de la tâche la plus complexe du test (donner un titre à chaque paragraphe, textes inconnus), les résultats me semblent probants et les réponses, bien que parfois incomplètes, étaient, dans la majeure partie des cas, fruits d'une réflexion poussée. La dernière partie du test, bien que décevante dans son ensemble, présente quelques points positifs : la première question et certaines réflexions quant à la seconde.

En conclusion, bien qu'il ne soit pas exempt de critiques, le test a bien mis en exergue les connaissances et les compétences des élèves. Tenant compte du fait qu'il s'agissait pour eux d'une « première » (jamais un test d'histoire n'avait été aussi « complexe ») et du contexte de travail (stage A, travail avec prafo, etc), les résultats sont probants.

## 2) Propositions de remédiations

Les observations menées ci-dessus permettent de dresser une liste de problèmes à remédier relativement aux activités d'enseignement-apprentissage. Comme il a été dit précédemment, la classe dans laquelle a été mené le test n'a aucunement l'habitude d'être confrontée à des tâches et des activités complexes en classe. C'est pourquoi j'ai mis en place plusieurs activités pour les faire travailler sur des documents authentiques, de manière autonome ou en groupes par exemple. Néanmoins, il semble évident que cela n'est pas suffisant : ce ne sont pas quelques périodes au cours du dernier semestre d'études qui ont permis de changer des habitudes bien ancrées.

Pour en revenir aux problèmes d'enseignement-apprentissage par rapport au test en question, les deux parties les plus complexes (D et E) me semblent devoir être retravaillées, les élèves n'ayant que trop peu l'habitude d'être confrontés à des tâches complexes impliquant des documents authentiques inconnus. C'est pourquoi il serait bon de revenir sur l'analyse de textes, leur compréhension et surtout la manière d'en tirer les informations principales et d'en faire une synthèse. Ce travail de longue haleine a été entraîné lors des cours, mais de manière trop succincte semble-t-il. Il conviendrait par exemple de mettre en place un travail de fond, avec grille d'analyse à l'appui, permettant aux élèves d'avoir un canevas précis, un exemple de marche à suivre pour l'analyse de textes historiques. En classe, nous avons fait pratiquement le même exercice demandé lors du test. Les élèves, répartis en groupes, devaient analyser une source (traitant de la doctrine émanant de *Mein Kampf*), les textes n'étant pas les mêmes que ceux du test : en soulignant les mots ou les passages clés, ils devaient présenter, dans un second temps, les idées principales de la source à leurs camarades et proposer un titre pour le passage étudié ; ce dernier était ensuite discuté avec le groupe. Malgré les bons résultats de ce travail de groupes, il est indéniable que ça n'est pas resté dans les mémoires et que les anciennes habitudes ont bien

vite repris le dessus. Il s'agit donc de répéter ce genre d'activités afin de donner des automatismes aux élèves.

Concernant la partie E et plus particulièrement la question 2, le faible taux de réussite de l'exercice prouve que l'analyse de graphique n'a pas bien été entraînée en classe. En effet, une analyse du même type a été menée en classe, mais de manière magistrale par l'enseignant, le temps faisant défaut pour mettre en place des activités socio-constructivistes. Il semble clair que la présentation menée en classe n'a pas été suffisante. Il serait bon de revenir sur les documents autres que les textes et d'en faire une analyse plus approfondie (un graphique en l'occurrence). Des activités de groupes où les élèves réfléchissent à des documents historiques (graphique, organigramme, tableaux, etc) pourraient mener à une meilleure consolidation des acquis et à une plus grande autonomie. Remarquons que, mis à part un ou deux problèmes, le travail de fond mené en classe visant à respecter les consignes et à lire attentivement les sources a porté ses fruits.

### 3) Analyse rétrospective du dispositif d'évaluation

Reprenant les observations menées dans la première partie de cette analyse, cinq points principaux sont à revoir : a) la pondération de la partie A (et peut-être C-2) ; b) la formulation de la consigne de l'exercice 3, partie B ; c) la question 1 de la partie C ; d) la difficulté relative du texte 1 de la partie D ; e) la consigne de la question 2, partie E. Ces cinq points sont à mettre en parallèle avec le taux élevé d'élèves sous le seuil de réussite dans le tableau ci-joint.

a) Il semble évident a posteriori que le nombre de points attribué à la partie A (connaissances pures) doit être pondéré. En effet, comme nous l'avons vu, un faible nombre de points porte à conséquence sur la note finale ; de plus, la configuration de l'exercice (replacer des événements historiques sur une frise) fait en sorte qu'une erreur implique forcément une autre. Il serait donc préférable de n'accorder qu'un demi point par réponse juste accordant ainsi 5 points totaux pour cette partie, ce qui serait plus en adéquation avec la complexité de la tâche et son impact sur la note finale.

La question 2 de la partie C comporte aussi 10 points (son poids sur la note finale est aussi conséquent) ; néanmoins il me semble que la complexité de la tâche demandée (analyse) est cohérente avec le nombre important de points attribués. Ainsi, je ne pense pas qu'il faille changer la pondération pour cette question.

b) La question 3 de la partie B pose problème vu le nombre relativement élevé d'élèves en dessous du seuil de réussite. Il est évident qu'il s'agit d'une question trop subjective. Bien que nous ayons parlé de divers traits de personnalités d'Hitler durant les cours (dont faisaient partie la paranoïa ou la paresse, d'où certaines réponses), je pense a posteriori que cette question n'a pas lieu d'être dans le test. La réponse peut paraître évidente pour

l'enseignant, mais les élèves n'ont pas forcément les mêmes références et le même vocabulaire... C'est pourquoi j'ai décidé d'accepter les réponses du type « racisme/agressivité » en donnant un demi point (j'attendais le mot « paranoïa »). Donc si je devais refaire ce test, cette question n'en ferait certainement plus partie, remplacée qu'elle serait par une autre plus objective.

c) Etrangement la question 1 de la partie C (connaissance pure) n'a pas été une réussite. Je n'arrive pas vraiment à me l'expliquer car le sigle NSDAP, sa signification, sa traduction française et l'abréviation qu'il a engendrée ont été vus et revus en classe... Peut-être aurait-il été bon d'écrire en entier le nom allemand du NSDAP, cela aurait peut-être éclairé certaines lanternes.

d) La complexité de la partie D du test a déjà été débattue longuement dans ce texte et il ne s'agit pas ici de savoir si une telle tâche peut ou non être testée (personnellement je pense que c'est un bon exercice). Par contre, le texte 1 est peut-être un peu trop compliqué par rapport aux activités effectuées en classe et au niveau du groupe par rapport à ce type de tâches complexes ; il s'agirait de trouver un autre exemple à analyser. La question de la tarification de l'exercice (2 points par texte) pourrait aussi être discutée. La correction ayant été faite avec mon praticien-formateur, nous avons décidé de pondérer les résultats par rapport à la pertinence de la réponse (analyse-synthèse) et au respect de la consigne (clarté et brièveté de la réponse) ; néanmoins quand une réponse était clairement hors sujet, aucun point n'était accordé. Il serait bon, dans le cas présent, de créer une grille de critères avec un certain nombre de points à attribuer le cas échéant.

e) Le problème relatif à la question 2 de la partie E, en plus de la relative complexité de la tâche, consiste dans la formulation de la consigne. En effet, sous indication de mon praticien-formateur, la première partie de l'énoncé qui mentionnait « En avril 1932 ... » donnant ainsi une réponse quant à la partie A a été enlevée. En résulte une consigne un brin amputée. Malgré ce petit bémol, le quasi totalité des élèves a compris de quelle élection il s'agissait et a su le replacer dans le contexte historique. N'ont manqué que les réflexions relatives aux conséquences de ces événements, mais cet état de fait a été discuté dans la partie précédente du texte présent. Quoi qu'il en soit, si c'était à refaire, je n'amputerai plus la consigne, quitte à transformer la partie A.

Tels sont les points à modifier quant au test. Il n'en reste pas moins que j'ai essayé de rester cohérent et d'appliquer au mieux mes connaissances théoriques sur « la réalité du terrain » ; ainsi ce test, sous la double contrainte (concrète : prafo, classe et théorique : didactique HEP), bien que particulièrement complexe à réaliser, a été amené d'une manière adéquate et que la grande majorité des élèves a prouvé avoir englobé de nouvelles connaissances et compétences.

#### 4) Bibliographie

La bibliographie présentée ici comporte de nombreux ouvrages portant sur l'évaluation de compétences en classe d'histoire. Au demeurant, je tenais à mentionner certains articles critiques quant à la logique de la compétence en éducation (Crahay (2006) ; Schneider-Gilot (2006) ; Meyer (1995). C'est en partie pour cela que mon test comporte diverses parties, faisant appel aussi bien à des compétences qu'à des connaissances.

Bouhon M. & Dambroise C. (dir.) (2002), *Evaluer des compétences en classe d'histoire*, Louvain-La-Neuve : Unité de didactique de l'Histoire de l'Université catholique de Louvain.

Bouhon M. & Jadoulle J.-L. & (2003), *Développer des compétences en classe d'histoire*, Louvain-La-Neuve : UCL.

Crahay M. (2006), "Dangers, incertitudes et incomplétude de la logique de la compétence en éducation", in *Revue française de pédagogie*, n° 154, janvier-février-mars 2006, pp. 97-110.

Gerard F.-M. (2005), «L'évaluation des compétences à travers des situations complexes», Actes du Colloque de l'Admee-Europe, IUFM Champagne-Ardenne, Reims, 24-26 octobre 2005

Jonnaert Ph. (2002) *Compétences et socioconstructivisme, un cadre théorique*, Bruxelles : De Boeck.

Meyer G. (1995), "Pratiquer l'évaluation normative en classe", in Meyer G., *Evaluer : pourquoi ? comment ?*, Paris : Hachette, pp. 38-42.

Roegiers X. (2003), *Des situations pour intégrer les acquis scolaires*, Bruxelles : De Boeck.

Rey B. & Staszewski M. (2004), *Enseigner l'histoire aux adolescents : Démarches socio-constructivistes*, Bruxelles : De Boeck.

Schneider-Gilot M. (2006), "Quand le courant pédagogique "des compétences" empêche une structuration des enseignements autour de l'étude et de la classification de questions parentes", in *Revue française de pédagogie*, n° 154, janvier-mars 2006, pp. 85-96