

Projet CLEF – Bénin

L'examen du CEP 2000 : analyse de l'instrumentation utilisée et des résultats obtenus

Rapport de mission de Léo Laroche

(22-01-01 au 02-02-01)

pour The Mitchell Group Inc.

Février 2001

Table des matières

| | |
|---|-----------|
| INTRODUCTION..... | 1 |
| RÉALISATION D'ACTIVITÉS PRÉPARATOIRES | 3 |
| COHÉRENCE ENTRE LA DÉFINITION DE DOMAINES ET LES TABLEAUX DE SPÉCIFICATION | 5 |
| PERTINENCE, FIABILITÉ ET VALIDITÉ DES ÉPREUVES..... | 5 |
| LES ÉPREUVES DU CEP, NOUVEAUX PROGRAMMES D'ÉTUDES | 5 |
| <i>L'épreuve d'expression écrite.....</i> | 5 |
| <i>L'épreuve de lecture.....</i> | 6 |
| <i>L'épreuve d'orthographe.....</i> | 6 |
| <i>L'épreuve de compréhension orale.....</i> | 6 |
| <i>L'épreuve de mathématique.....</i> | 6 |
| <i>L'épreuve d'éducation sociale et d'éducation scientifique et technologique.....</i> | 7 |
| <i>L'épreuve de dessin.....</i> | 7 |
| <i>L'épreuve d'éducation physique et sportive</i> | 7 |
| LES ÉPREUVES DU CEP, PROGRAMMES INTERMÉDIAIRES..... | 8 |
| <i>La dictée.....</i> | 8 |
| <i>La rédaction.....</i> | 8 |
| <i>L'interrogation écrite.....</i> | 8 |
| <i>L'étude de texte.....</i> | 9 |
| <i>Le calcul.....</i> | 9 |
| <i>Le dessin ou la couture.....</i> | 9 |
| CONCLUSIONS..... | 9 |
| ANALYSE STATISTIQUE ET PÉDAGOGIQUE DES RÉSULTATS OBTENUS..... | 10 |
| LE FRANÇAIS..... | 11 |
| <i>L'expression écrite.....</i> | 11 |
| <i>La lecture.....</i> | 12 |
| <i>L'orthographe.....</i> | 13 |
| LA MATHÉMATIQUE..... | 14 |
| L'ÉDUCATION SCIENTIFIQUE ET TECHNIQUE / L'ÉDUCATION SOCIALE..... | 16 |
| L'ÉDUCATION ARTISTIQUE..... | 21 |
| L'ÉDUCATION PHYSIQUE ET SPORTIVE | 22 |
| COMPARAISONS DES RÉSULTATS OBTENUS AU CEP, NOUVEAUX PROGRAMMES, ET AU CEP, PROGRAMMES INTERMÉDIAIRES..... | 23 |
| COÛT ET EFFICACITÉ DU NOUVEAU FORMAT DES ÉPREUVES..... | 24 |
| LA SESSION 2000 DU CEP, NOUVEAUX PROGRAMMES..... | 24 |
| ACTIVITÉS COMPLÉMENTAIRES RÉALISÉES AU COURS DE LA MISSION | 24 |
| ACTIVITÉS QUOTIDIENNES..... | 25 |
| LE LUNDI, 22 JANVIER 2001..... | 25 |
| LE MARDI, 23 JANVIER 2001..... | 26 |
| LE MERCREDI, 24 JANVIER 2001..... | 26 |
| LE JEUDI, 25 JANVIER 2001..... | 26 |
| LE VENDREDI, 26 JANVIER 2001..... | 27 |
| LE SAMEDI, 27 JANVIER 2001..... | 27 |
| LE LUNDI, 29 JANVIER 2001..... | 28 |
| LE MARDI, 30 JANVIER 2001..... | 28 |

| | |
|---|-----------|
| LE MERCREDI, 31 JANVIER 2001..... | 28 |
| LE JEUDI, 1 ^{ER} FÉVRIER 2001 | 28 |
| LE VENDREDI, 2 FÉVRIER 2001..... | 29 |
| SUITE DES OPÉRATIONS | 29 |
| ANNEXE « A » QUELQUES DÉFINITIONS DE CONCEPTS RELIÉS À L'ÉVALUATION..... | 31 |
| ANNEXE « B » TABLEAUX DE SPÉCIFICATION RÉVISÉS..... | 38 |
| ANNEXE « C » RAPPORT SOMMAIRE..... | 47 |

Introduction

La présente mission au Bénin est la suite de séjours réalisés en 1999 et 2000. Ces précédentes missions ont touché : a) le traitement et l'analyse des résultats de l'examen de juillet 1998 du CEP, nouveaux programmes d'études; b) la préparation de jeux d'épreuves destinées aux sessions de juin 1999 et 2000 de l'examen du CEP; c) réalisation d'une analyse statistique des épreuves administrées en juillet 1999; d) examen du format de CEP, nouveaux programmes d'études et proposition d'un format modifié à être utilisé en juin 2000. Ce rapport touche la mission de **Léo Laroche**, ci-après nommé **le consultant**, qui s'est déroulée du 22 janvier au 2 février 2001. Voici un rappel des termes de référence de cette mission.

L'objet de la présente mission a ainsi été défini par les responsables béninois :

- *Evaluation des résultats du CEP nouveaux programmes d'études, session de Juin 2000.*

Le contexte de l'intervention¹ ainsi que sa justification ont été ainsi formulés :

- *Dans le cadre de la généralisation prochaine du nouveau format des épreuves du CEP nouveaux programmes d'études, il est nécessaire d'une part de procéder à l'évaluation des résultats du CEP 2000 administré à partir d'un format en expérimentation et d'autre part de renforcer les capacités nationales en matière de conception, d'assemblage, d'administration des épreuves, de correction et d'analyse des résultats à travers la mise en place d'instruments appropriés (tableaux de bord...) qui facilitent la gestion efficiente de banque d'items et d'épreuves.*

Voici donc les objectifs de la consultation :

- *En collaboration avec les membres de l'équipe du Plan d'action Evaluation des apprentissages et Orientation Scolaire, les cadres de la Direction des Examens et Concours et les assistantes techniques du Projet CLEF, le consultant devra :*
 1. *déterminer le degré de cohérence entre la définition de domaine des programmes CM₂ et le tableau de spécification des épreuves proposées ;*
 2. *examiner la pertinence, la fiabilité et la validité des épreuves proposées dans chaque champ de formation ;*
 3. *faire une analyse statistique et pédagogique des résultats obtenus au CEP 2000 par les candidats ;*
 4. *évaluer le rapport coût - efficacité du nouveau format des épreuves du CEP, nouveaux programmes ;*
 5. *faire une évaluation globale de cette session et proposer des recommandations pour l'amélioration du contenu, de l'administration et de la correction des prochaines sessions du CEP, nouveaux programmes d'études.*

¹ La durée de la consultation devait originellement s'étendre sur trois semaines. Ramenée à deux semaines, cette mission n'a pas pu comprendre les volets reliés à la formation du personnel par rapport à la constitution de banques d'items mises en place pour la préparation des futures sessions du CEP, nouveaux programmes d'études.

Les résultats attendus ont été ainsi exprimés :

- *le degré de cohérence entre la définition de domaine des programmes CM₂ et le tableau de spécification des épreuves proposées dans chaque champ de formation aura été déterminé ;*
- *l'examen de la pertinence, la fiabilité et la validité des épreuves proposées dans chaque champ de formation aura été effectué ;*
- *l'analyse statistique et pédagogique des résultats obtenus par les candidats au cours du CEP 2000 aura été faite ;*
- *l'évaluation du rapport coût-efficacité du nouveau format des épreuves du CEP, nouveaux programmes aura été faite ;*
- *un rapport détaillé portant sur l'évaluation des résultats du CEP, nouveaux programmes d'études, session de Juin 2000, rapport assorti de propositions concrètes pour l'amélioration des prochaines sessions du CEP aura été rédigé.*

Le présent rapport se compose des chapitres suivants :

- Réalisation d'activités préparatoires.
- Cohérence entre la définition de domaines et les tableaux de spécification.
- Pertinence, fiabilité et validité des épreuves.
- Analyse statistique et pédagogique des résultats obtenus.
- Coût et efficacité du nouveau format des épreuves.
- Évaluation de la dernière session du CEP, nouveaux programmes.
- Rappel des activités quotidiennes.
- Bilan et suite des opérations.

Des annexes fournissent des renseignements complémentaires.

Réalisation d'activités préparatoires

Le consultant a reçu, avant son départ pour le Bénin, une série de fichiers contenant des échantillons de notations à sept épreuves² administrées en juin 2000 dans le cadre du CEP, nouveaux programmes d'études. Les bases de données ont été examinées; il y a eu production de statistiques pour examiner les qualités docimologiques de ces instruments. Voici un sommaire établi à la suite de l'examen des résultats obtenus³.

Tableau 1
Résultats sommaires aux épreuves de juin 2000

| Épreuves | Nbre éléments | Échantillon | Moyenne/100 | Consistance | Discrimination | Taux de succès |
|-------------------|---------------|-------------|-------------|-------------|----------------|----------------|
| EST/ES | 31 | 117 | 44,8 | 0,830 | 25,8 | 69,2 |
| EPS | 5 | 100 | 79,5 | 0,880 | 100,0 | 94,0 |
| Lecture | 7 | 118 | 65,5 | 0,470 | 57,1 | 89,8 |
| Expression écrite | 10 | 120 | 45,9 | 0,350 | 40,0 | 72,5 |
| Orthographe | 1 | 119 | 62,0 | | | 76,5 |
| Mathématique | 18 | 115 | 52,0 | 0,830 | 77,8 | 65,2 |
| Dessin | | | | | | |

Voici les critères retenus pour déterminer le degré de discrimination des épreuves et le taux de succès des élèves. Pour ce qui est la discrimination, nous avons retenu un coefficient de corrélation de 0,400 ou supérieur ; quant au taux de succès, nous avons retenu la note 50 sur 100 comme limite pour indiquer que l'élève a réussi à l'examen.

Nous pouvons faire les commentaires suivants à la suite de l'examen des statistiques contenues dans le tableau numéro 1 :

- L'épreuve d'*éducation scientifique et technologique / d'éducation sociale* comprend un nombre suffisant d'éléments d'information pour couvrir les disciplines. La moyenne obtenue par les 117 élèves faisant partie de l'échantillon est faible ; ceci affecte le taux de succès à l'épreuve (69%). La consistance interne est acceptable ; toutefois, la discrimination des différents items contenus dans l'épreuve est plutôt faible dans 75% des cas.
- L'épreuve d'*éducation physique et sportive* a été réussie par 94% des élèves faisant partie de l'échantillon retenu. La consistance interne ainsi que la discrimination sont acceptables. On doit noter qu'il n'y a que cinq critères retenus pour porter le jugement concernant l'atteinte des objectifs attendus.

² Il n'a pas été possible de traiter les données portant sur l'épreuve de dessin; il y avait alors une incohérence entre la base de données et les fichiers de saisie de l'information.

³ Comme nous le verrons plus loin, les bases de données ont dû être révisées au cours de la mission de consultation. Des enregistrements ont été éliminés (absence de ces élèves à certaines épreuves), les codes attribués à certaines variables furent corrigés. Ceci a affecté légèrement certaines statistiques. Nous les fournissons afin de mieux situer le contexte des travaux réalisés dès l'arrivée du consultant à Pobè.

- Près de 90% des élèves ont réussi l'épreuve de *lecture*. La consistance interne de l'épreuve est très faible ; il faudra examiner cette question avec les spécialistes de la discipline. Près de la moitié des items faisant partie de l'épreuve discrimine bien, les autres ayant un coefficient plutôt faible. On peut noter que seulement sept éléments d'information sont contenus dans cette épreuve.
- Dix critères sont utilisés pour la correction de l'épreuve d'*expression écrite*. La moyenne obtenue est assez faible (moins de 46 sur 100) ; ceci affecte le taux de succès à l'épreuve qui se situe à 72%. Toutefois, ce sont les indices de discrimination et de consistance interne qui doivent préoccuper les responsables de cette épreuve ; dans les deux cas, ces indices sont faibles, indiquant que l'épreuve ne respecte pas tous les critères caractérisant une épreuve de qualité.
- L'épreuve d'*orthographe* ne contient qu'un item. Il n'est donc pas possible d'en faire une analyse utilisant des coefficients produits à la suite de l'analyse d'items. On peut toutefois indiquer que plus de 75% des élèves de l'échantillon considéré ont réussi l'épreuve.
- L'épreuve de *mathématique* comprend un nombre suffisant d'éléments d'information permettant sans doute de bien couvrir le tableau de spécification. Les indices de consistance interne et de discrimination sont élevés indiquant que l'épreuve respecte les standards ordinairement attendus. Le taux de succès ainsi que la moyenne générale sont plutôt faibles ; il s'agit d'une épreuve de difficulté assez grande pour les élèves évalués.
- Comme nous l'indiquions plus haut, il n'a pas été possible de faire l'analyse de l'épreuve de *dessin* à partir des renseignements reçus avant l'arrivée au Bénin.

On peut indiquer aussi quelques commentaires d'ordre générale :

1. Les bases de données ne contiennent pas de liens entre les éléments d'information et les tableaux de spécification⁴.
2. Le nombre d'élèves faisant partie des échantillons retenus est très faible. Il ne faudrait pas vouloir généraliser les résultats à partir des données actuellement disponibles. Il s'agit plutôt d'un exercice fournissant des indications utiles aux travaux qui seront réalisés à la suite de l'administration de la session de juin 2000 du CEP, nouveaux programmes.
3. Il faudra examiner avec plus d'attention les trois épreuves portant sur le français. Il est sans doute possible d'apporter des modifications susceptibles d'améliorer les qualités métriques de ces épreuves.
4. Toutes les questions posées aux élèves dans le cadre de ces sept épreuves sont du type « à crédits partiels ». Il faudrait examiner la possibilité d'introduire d'autres types d'items qui ne nécessiteraient pas nécessairement une notation requérant un jugement de la part des correcteurs.

⁴ Cette situation a été corrigée au cours du séjour du consultant.

Cohérence entre la définition de domaines et les tableaux de spécification

Avant l'arrivée du consultant au Bénin, il y a eu des activités réalisées par les responsables béninois touchant l'examen de l'instrumentation utilisée en juin 2000 par rapport aux différents tableaux de spécification élaborés en mars 2000. On y a constaté que certaines épreuves ne respectent pas complètement les pondérations comprises dans ces tableaux. Il faut noter qu'il peut être difficile de respecter strictement la répartition des notes à attribuer en se référant à une pondération établie par rapport au programme d'études. Ceci est d'autant plus valable lorsque le nombre d'éléments d'information recueillis par l'épreuve est plutôt faible; et il en est ainsi pour plusieurs épreuves administrées en juin 2000 dans le cadre du CEP, nouveaux programmes. Pour être en mesure de respecter plus rigoureusement les paramètres fixés dans les tableaux de spécification, il faudrait donc s'assurer que les épreuves recueillent davantage d'éléments d'information; la couverture des champs de formation par les épreuves serait alors probablement plus complète. Cependant, une telle façon de procéder aurait un impact sur le format des épreuves et sur les coûts reliés à leur administration. Il importe alors de faire les choix qui respectent tant la rigueur de l'évaluation que les capacités d'administration du mécanisme.

À la suite de l'analyse de l'instrumentation utilisée en juin 2000, les tableaux de spécification furent donc révisés pour mieux tenir compte des programmes d'études; l'annexe « B » contient ces nouveaux tableaux de spécification. Ils serviront au moment de l'élaboration des épreuves devant être utilisées en juin 2001.

Pertinence, fiabilité et validité des épreuves

Cette section du rapport passe en revue l'instrumentation utilisée en juin 2000 pour l'administration du CEP. Dans un premier temps, nous examinerons chaque épreuve destinée aux élèves ayant suivi les nouveaux programmes d'études. Par la suite, il sera question du matériel prévu pour la majorité des candidats, c'est-à-dire ceux ayant suivi les programmes intermédiaires. Il faut noter que les participants à l'atelier de Pobè ont fait une analyse rigoureuse de chaque épreuve administrée aux élèves fréquentant les classes des écoles expérimentales; il faudrait se référer aux rapports produits par les différentes équipes de travail.

Les épreuves du CEP, nouveaux programmes d'études

Huit épreuves sont prévues pour l'examen du CEP, nouveaux programmes. Il y a eu quatre épreuves de français (expression écrite, lecture, orthographe, compréhension orale), une épreuve de mathématique, une épreuve de dessin, une épreuve d'éducation physique et sportive et une épreuve combinant les deux champs de l'éducation sociale et de l'éducation scientifique et technologique.

L'épreuve d'expression écrite

La grille de correction contient des critères permettant d'apprécier un projet d'écriture. La plupart des critères (il y en a dix) sont notés de façon binaire («respect» / «non-respect»). Il pourrait y avoir une échelle descriptive comportant trois ou quatre points d'ancrage. Cela faciliterait la notation et rendrait plus uniforme l'appréciation à l'ensemble des critères retenus.

L'épreuve de lecture

Cette épreuve suppose, dans un premier temps, que les élèves lisent un texte décrivant un match de football. Le texte est assez long (plus de vingt lignes). Quatre paragraphes forment le texte. Le premier paragraphe n'est constitué que d'une phrase, le deuxième occupant plus de la moitié de la longueur totale du texte. Le récit occupant la première page de l'épreuve, il est nécessaire pour l'élève de tourner et retourner les pages pour chercher l'information et répondre aux questions.

La grille de correction prévue présente la répartition des notes dans un format laissant supposer qu'il s'agit d'attribuer les points sur un continuum, c'est-à-dire avec possibilité de «crédits partiels»; dans la majorité des situations, il s'agit plutôt de notation de type «correcte» ou «incorrecte».

La première question posée aux élèves aurait pu être plus précise en indiquant le nombre attendu de phrases respectant le critère fixé.

L'épreuve d'orthographe

On semble avoir demandé aux correcteurs de fournir uniquement une note globale à cette épreuve qui est intégrée à l'expression écrite. Il faudrait plutôt rapporter le nombre de fautes des deux catégories retenues, à savoir l'orthographe d'usage et l'orthographe grammaticale. Plutôt que d'indiquer qu'une faute d'orthographe grammaticale vaut deux fautes, il serait préférable de noter le nombre de points perdus par catégorie de fautes. Il est étonnant que la distribution de fréquences des notes obtenues par les élèves soient exprimées uniquement par des nombres pairs; le tableau de conversion pour établir la note obtenue ne prévoit que des nombres pairs. Il serait préférable de définir un tel tableau permettant la présence de toutes les notes situées entre 0 et 10.

L'épreuve de compréhension orale

Tout comme pour l'épreuve d'expression écrite, il serait préférable de préparer une véritable grille de notation comportant des indicateurs décrivant chaque point de l'échelle d'appréciation. Au moment de la notation, il serait pertinent d'indiquer le numéro de la question qui a été choisie. La décision de fournir à l'élève des éléments d'information après la première question si elle a été incorrectement répondue est appropriée; ceci enlève la possibilité d'avoir une dépendance locale entre la première question posée et les autres.

L'épreuve de mathématique

L'épreuve de mathématique suppose que l'élève fournisse ses réponses dans un cahier différent des textes constituant l'interrogation proprement dite. Plusieurs questions sont du type «correcte» / «incorrecte» alors que la grille de correction laisse croire que la correction est plutôt de type «à crédits partiels». Il faudrait réviser l'épreuve en conséquence lors de l'élaboration de l'épreuve de juin 2001.

L'épreuve d'éducation sociale et d'éducation scientifique et technologique

Cette épreuve comprend neuf questions touchant les deux champs de formation. On peut tout d'abord s'interroger sur la pertinence d'utiliser des chiffres romains pour désigner les questions. Par ailleurs, on pourrait uniformiser la numérotation des éléments constituant chaque question (par exemple, utiliser des lettres en minuscule). Autre commentaire par rapport à la présentation : il faudrait éviter de fractionner une question en la plaçant sur deux pages différentes (c'est le cas de la question numéro 5).

L'élève fournit ses réponses à cette épreuve dans un cahier prévu à cet effet. Tout comme plusieurs autres épreuves, certaines questions sont plutôt du type «correct» / «incorrect»; il faudrait que la grille de correction soit préparée pour tenir compte de cette réalité.

La question numéro 8 contient une liste de quatre raisons expliquant l'arrivée des Français au Bénin. On demande de recopier les raisons véritables. Il serait préférable d'indiquer à l'élève le nombre de raisons que l'on attend. La clé de correction est très complexe puisque les combinaisons sont nombreuses. Il faudrait tenter de trouver une autre façon de poser ce genre de questions.

La question numéro 9 comporte les mêmes caractéristiques que la question 8. Il faudrait faire des modifications semblables dans ce cas-ci.

L'épreuve de dessin

La grille de correction comporte des échelles à cinq points d'ancrage. Le dernier critère ne comprend que quatre points, le minimum de la notation étant 1 (le zéro est absent). Il serait préférable d'avoir le même nombre de points à chaque critère.

La grille pourrait être disposée autrement, en plaçant les indicateurs dans les cellules plutôt qu'à gauche du document.

L'épreuve d'éducation physique et sportive

Le matériel disponible pour assurer l'administration de cette épreuve est complet et semble adapté aux objectifs poursuivis. Il serait possible d'utiliser le même matériel d'une année à l'autre. Le choix des composantes offre plusieurs combinaisons admissibles.

Les épreuves du CEP, programmes intermédiaires

Nous avons pu examiner les épreuves qui ont été utilisées pour l'examen du CEP, programmes intermédiaires. L'instrumentation comprend six épreuves :

- La dictée
- La rédaction
- L'étude de texte
- Le calcul
- Le dessin ou la couture
- L'interrogation écrite

En français, l'orthographe est évaluée à l'aide d'une dictée. Il ne semble pas y avoir d'épreuve d'éducation physique. Nous ne disposons pas des grilles de correction pour en faire l'analyse; aussi, nous faisons porter nos commentaires uniquement sur les instruments remis aux candidats.

La dictée

Cette épreuve est constituée d'un texte comprenant environ sept lignes réparties dans trois paragraphes. Les élèves devaient écrire le texte lu sans faire de fautes d'orthographe grammaticale ou d'orthographe d'usage.

La rédaction

Le sujet de la rédaction porte sur la vaccination. Après une courte mise en situation (une phrase indiquant sur quoi porte la rédaction), la consigne est très brève. On demande à l'élève de raconter. On ne fournit pas d'indications supplémentaires (par exemple, la longueur attendue).

L'interrogation écrite

Ce que l'on désigne comme étant l'interrogation écrite touche les champs de formation Éducation sociale et Éducation scientifique et technologique. Six questions sont posées. Certaines de ces questions exigent de fournir des renseignements sur plusieurs objets (par exemple, la question numéro 5 comprend la reproduction d'une carte du Bénin et la localisation de quatre ressources minières).

L'étude de texte

Un texte comprenant moins de quinze lignes est proposé aux élèves. Après la lecture du texte, le candidat doit répondre à des interrogations et accomplir des tâches touchant la conjugaison et la grammaire. L'épreuve comporte des questions et tâches par rapport aux aspects suivants :

- Le vocabulaire
- La conjugaison
- La grammaire
- L'intelligence du texte.

Le calcul

Les questions posées aux candidats dans cette épreuves sont regroupés par rapport aux catégories suivantes :

- Les opérations
- Les exercices
- Un problème.

Le dessin ou la couture

Pour cette épreuve, le candidat choisit entre la tâche proposée en dessin (dessiner un moyen de déplacement et l'identifier) et la couture (fabrication d'une boutonnière).

Conclusions

Nous avons pu noter quelques faiblesses par rapport aux épreuves utilisées en juin 2000 dans le cadre du CEP, nouveaux programmes. Les lacunes identifiées touchent parfois les épreuve elles-mêmes, souvent les grilles de correction. La présentation des *épreuves* remises aux élèves est soignée et il n'y a pas d'erreurs pouvant avoir une influence significative sur les réponses fournies par les élèves. Il y aurait toutefois lieu de mieux standardiser les *consignes* là où cela s'applique. Dans cet ordre d'idées, les Guides d'administration pourraient être intégrés, devenant alors un document contenant l'ensemble de l'information utile pour assurer une gestion harmonieuse de l'examen; un tel document n'aurait probablement pas à être repris d'une année à l'autre, à moins que des modifications aux procédures aient été faites. Là où il y aurait des changements à apporter concerne les *guides de correction*. L'ambiguïté de certaines consignes de notation a certainement eu un impact sur les résultats obtenus par les élèves. Il serait souhaitable d'uniformiser certains guides de correction, en particulier là où il y a des critères d'évaluation requérant l'attribution de notes suivant la technique de «crédits partiels».

Comme nous l'avons indiqué aux participants de l'atelier de Pobè, on peut classer les épreuves administrées en juin 2000 par rapport à cette classification : a) des épreuves écrites recueillant des éléments d'information fournis par des réponses des élèves à une série de questions; b) des tâches que l'on demande de réaliser et dont la production est évaluée à l'aide d'une grille de notation ou d'un schéma d'observation. Ordinairement, la première catégorie citée comprend des épreuves dont l'information peut

être traitée plus facilement; on peut rattacher les épreuves suivantes à ce type d'épreuves : mathématique, éducation scientifique et technologique, éducation sociale, lecture. Les autres épreuves du CEP peuvent être considérées comme des tâches à réaliser par les élèves et dont la correction sera réalisée à l'aide d'une série de critères ou d'indicateurs de réussite. C'est ici qu'il faudrait concentrer ses efforts pour d'une part mieux préciser les critères à considérer et d'autre part assurer une meilleure formation aux correcteurs à l'utilisation d'un tel outil d'évaluation.

Analyse statistique et pédagogique des résultats obtenus

Comme nous l'indiquons plus haut, des données ont été saisies sur un échantillon de copies d'élèves soumis aux épreuves du CEP, nouveaux programmes, de juin dernier. Environ vingt élèves par département ont été retenus; ce sont les mêmes élèves qui ont été choisis pour l'ensemble des épreuves. L'objectif visé par cet exercice est d'examiner la qualité métrique de ces instruments. Il est vrai que la taille de l'échantillon aurait pu être plus grande; les données disponibles permettront quant même de disposer de tendances susceptibles de fournir des renseignements utiles pour se prononcer sur la qualité des épreuves utilisées et guider les concepteurs des futures épreuves. Nous retiendrons deux types de compilations statistiques : dans un premier temps, il y a eu analyse d'items afin d'examiner les qualités métriques des épreuves; par la suite, le calcul de résultats moyens par regroupement d'items correspondant aux tableaux de spécification permet d'identifier les habilités et les contenus notionnels plus ou moins bien réussis. Nous fournissons des statistiques sur chacune des épreuves⁵; un premier tableau permet de disposer d'un sommaire d'information sur chaque épreuve.

Tableau 2
Renseignements statistiques sommaires sur les épreuves de juin 2000

| Épreuves | N | Éléments | Moyenne | Succès | Alpha | Difficiles | Peu discriminants |
|--------------|-----|----------|---------|--------|-------|------------|-------------------|
| Mathématique | 115 | 18 | 52,1 | 53,9 | 0,834 | 6 | 4 |
| EST / ES | 117 | 31 | 44,9 | 47,9 | 0,831 | 19 | 15 |
| Exp. Écrite | 117 | 10 | 44,7 | 41,7 | 0,380 | 5 | 4 |
| Lecture | 117 | 7 | 66,1 | 87,2 | 0,419 | 0 | 1 |
| Orthographe | 117 | 1 | 63,1 | 77,8 | ----- | ----- | ----- |
| Dessin | 117 | 4 | 76,4 | 100,0 | 0,593 | 0 | 0 |
| EPS | 94 | 5 | 84,2 | 100,0 | 0,512 | 0 | 0 |

À la lecture de ce tableau, on peut observer que le nombre d'éléments d'information présents dans les différentes épreuves varie considérablement (de un élément à 31). Il est entendu que plus il y a d'éléments d'information disponibles, plus fiables seront les indices statistiques rendus disponibles par des compilations statistiques. Certaines épreuves ont été très difficiles pour les élèves (Expression écrite, EST / ES, Mathématique); d'autres (EPS et Dessin) ont été très faciles. La cohérence interne des épreuves est

⁵ On doit noter que l'épreuve de communication orale n'a pas pu être traitée.

acceptable là où l'on dispose d'un nombre suffisant d'éléments d'information (mathématique et éducation scientifique et technologique / éducation sociale). Quant à la discrimination des différents items ou critères, il y a généralement des lacunes qui mériteraient d'être examinées avec attention. Passerons immédiatement en revue les résultats obtenus à chaque épreuve.

Le français

Quatre épreuves portant sur le français font partie de l'examen du CEP de juin 2000, se répartissant 85 points dans le calcul de la note globale pour l'obtention de ce certificat :

- Lecture 30 points
- Expression écrite 25 points
- Orthographe 10 points
- Communication orale 20 points

On doit noter ici que l'orthographe est évaluée à partir de l'épreuve d'Expression écrite. Au moment de la tenue de l'atelier, nous ne disposons pas de données sur l'épreuve de communication orale.

L'expression écrite

La production écrite a été corrigée par rapport à quatre critères :

- L'organisation des idées
- Le choix des idées
- La formulation des idées et la syntaxe
- La calligraphie et la présentation

La majorité de ces critères comporte des indicateurs qui ont servi à réaliser la correction. Il y a ainsi dix éléments d'information qui permettent de rendre compte de la performance des élèves évalués. Le tableau 3 reproduit les statistiques obtenues.

Tableau 3
Résultats obtenus à l'épreuve d'expression écrite

| Items | N= | Minimum | Maximum | Moyenne | Rés/100 | Corr. |
|-------|-----|---------|---------|---------|---------|--------|
| Q1A | 115 | 0,00 | 2,00 | 1,53 | 76,52 | 0,280 |
| Q1B | 115 | 0,00 | 3,00 | 2,44 | 81,45 | 0,597 |
| Q1C1 | 115 | 0,00 | 1,00 | 0,56 | 55,65 | 0,195 |
| Q1C2 | 115 | 0,00 | 1,00 | 0,70 | 70,43 | -0,011 |
| Q2A | 115 | 0,00 | 4,00 | 1,35 | 33,70 | 0,504 |
| Q2B | 115 | 0,00 | 3,00 | 1,69 | 56,23 | 0,503 |
| Q2C | 115 | 0,00 | 4,00 | 0,84 | 21,09 | 0,277 |
| Q3A | 115 | 0,00 | 3,00 | 0,64 | 21,30 | 0,374 |
| Q3B | 115 | 0,00 | 2,00 | 0,76 | 37,83 | 0,568 |
| Q4 | 115 | 0,00 | 2,00 | 0,67 | 33,70 | 0,377 |

| Résultats | Moyenne | Ecart type | Alpha | Err-type de mesure |
|---------------|---------|------------|-------|--------------------|
| <i>Global</i> | 44,73 | 15,41 | 0,38 | 12,13 |

Le résultat moyen de cette épreuve s'établit à 45 sur 100. La moitié des indicateurs utilisés pour corriger la production écrite a obtenu un indice de difficulté élevé, c'est-à-dire que leur résultat moyen est très faible (entre 21 et 34 sur 100). En ce qui concerne la discrimination mesurée par la corrélation item – résultat global obtenu à l'épreuve par chaque élève, le premier critère de correction relatif à l'organisation des idées est plutôt faible.

On doit signaler une valeur assez faible de l'indice Alpha; il y aurait peu de liens entre les différents éléments d'information recueillis à la suite de la correction de la production écrite.

La lecture

L'épreuve de Lecture comprend sept éléments d'information répartis dans quatre questions posées aux élèves. Le tableau 4 présente les résultats obtenus à cette épreuve.

Tableau 4
Résultats obtenus à l'épreuve de lecture

| Items | N= | Minimum | Maximum | Moyenne | Rés/100 | Corr. |
|-------|-----|---------|---------|---------|---------|-------|
| Q1A | 117 | 0,00 | 7,50 | 3,58 | 47,69 | 0,327 |
| Q1B1 | 117 | 0,00 | 0,50 | 0,43 | 86,32 | 0,334 |
| Q1B2 | 117 | 0,00 | 2,00 | 1,42 | 70,94 | 0,470 |
| Q2A | 117 | 0,00 | 2,00 | 1,35 | 67,52 | 0,197 |
| Q2B | 117 | 0,00 | 2,00 | 1,15 | 57,26 | 0,392 |
| Q3 | 117 | 0,00 | 12,00 | 8,83 | 73,58 | 0,775 |
| Q4 | 117 | 0,00 | 4,00 | 3,07 | 76,71 | 0,583 |

| Résultats | Moyenne | Ecart type | Alpha | Err-type de mesure |
|----------------------|---------|------------|-------|--------------------|
| <i>Global</i> | 66,07 | 18,98 | 0,42 | 14,47 |
| <i>Les habiletés</i> | | | | |
| INFORMATION | 54,27 | 22,31 | 0,17 | 20,26 |
| SENS | 73,58 | 32,56 | | |
| INFERENCE | 62,39 | 42,88 | 0,76 | 20,90 |
| CODE | 76,71 | 41,75 | | |

Le résultat moyen de l'épreuve s'établit à 19 sur 30 ou à 66,1 sur 100. Il s'agit d'une épreuve plutôt facile. Deux éléments d'information (la question 1A relative à l'identification de phrases indiquant les sentiments des spectateurs et la question 2B demandant de justifier une réponse fournie) ont obtenu des résultats plutôt faibles. Quant à la discrimination, un seul item comporte un coefficient plutôt faible (la question 2A); l'élève devait faire une inférence en marquant son choix sur une interrogation de type «Vrai / Faux». La consistance interne de l'épreuve mesurée par le coefficient Alpha est plutôt faible; on doit noter ici que le nombre peu élevé d'éléments d'information explique en partie la valeur obtenue à cet indice statistique.

Quatre habiletés regroupent les items de l'épreuve. Les habiletés reliées au respect du code et le sens des mots sont mieux réussies que celles touchant la sélection de l'information et l'inférence.

L'orthographe

L'information codifiée pour rendre compte de la maîtrise de l'orthographe est globale; seulement la note sur 10 attribuée aux élèves est disponible dans la base de données créée pour cette activité d'analyse. On peut donc uniquement rendre compte du résultat global obtenu par les élèves faisant partie de l'échantillon.

Tableau 5
Résultats obtenus à l'épreuve d'orthographe

| Items | N= | Minimum | Maximum | Moyenne | Rés/100 | Err-type de mesure |
|---------------|-----------|----------------|----------------|----------------|----------------|-------------------------------|
| <i>Global</i> | 117 | 2,00 | 10,00 | 6,31 | 63,08 | 20,45 |

Ce résultat moyen, qui s'établit à 63 sur 100, a étonné les participants à l'atelier; on croit que ce résultat ne rend pas compte de la réussite réelle des élèves dans cette discipline; les élèves béninois ne sembleraient pas maîtriser aussi complètement l'orthographe d'usage et grammaticale. Il aurait fallu disposer d'une grille de correction plus précise pour être en mesure d'établir un meilleur diagnostic en ce qui concerne la performance des élèves en orthographe.

La mathématique

Cette épreuve comprend six problèmes touchant les contenus notionnels et les habiletés faisant partie du tableau de spécification. Chaque problème comporte trois éléments d'information pour couvrir les différentes habiletés mesurées.

Tableau 6
Résultats obtenus à l'épreuve de mathématique

| Items | N= | Minimum | Maximum | Moyenne | Rés/100 | Corr. |
|-------|-----|---------|---------|---------|---------|-------|
| Q1A | 115 | 0,00 | 3,00 | 2,15 | 71,59 | 0,586 |
| Q1B | 115 | 0,00 | 3,00 | 1,77 | 59,13 | 0,553 |
| Q1C | 115 | 0,00 | 3,00 | 1,73 | 57,68 | 0,558 |
| Q2A1 | 115 | 0,00 | 5,00 | 3,37 | 67,30 | 0,613 |
| Q2A2 | 115 | 0,00 | 4,00 | 2,55 | 63,70 | 0,640 |
| Q2B | 115 | 0,00 | 3,00 | 1,43 | 47,54 | 0,550 |
| Q3A1 | 115 | 0,00 | 2,00 | 1,38 | 69,13 | 0,495 |
| Q3A2 | 115 | 0,00 | 2,00 | 1,04 | 52,17 | 0,645 |
| Q3B | 115 | 0,00 | 4,00 | 1,40 | 35,00 | 0,660 |
| Q4A | 115 | 0,00 | 3,00 | 2,11 | 70,43 | 0,532 |
| Q4B1 | 115 | 0,00 | 3,00 | 1,61 | 53,62 | 0,669 |
| Q4B2 | 115 | 0,00 | 4,00 | 1,25 | 31,30 | 0,651 |
| Q5A | 115 | 0,00 | 3,00 | 2,25 | 75,07 | 0,231 |
| Q5B | 115 | 0,00 | 3,00 | 1,83 | 61,16 | 0,299 |
| Q5C | 115 | 0,00 | 3,00 | 0,91 | 30,43 | 0,451 |
| Q6A | 115 | 0,00 | 4,00 | 1,57 | 39,13 | 0,209 |
| Q6B | 115 | 0,00 | 4,00 | 2,05 | 51,30 | 0,298 |
| Q6C | 115 | 0,00 | 4,00 | 0,81 | 20,22 | 0,493 |

| Résultats | Moyenne | Ecart type | Alpha | Err-type de mesure |
|--------------------------|---------|------------|-------|--------------------|
| <i>Global</i> | 52,03 | 21,74 | 0,83 | 8,86 |
| <i>Habilités</i> | | | | |
| STRUCTURATION | 59,48 | 28,23 | 0,64 | 16,93 |
| OPERATION | 45,22 | 26,00 | 0,54 | 17,63 |
| ANALYSE | 50,48 | 27,33 | 0,85 | 10,72 |
| <i>Contenu notionnel</i> | | | | |
| ARITHMETIQUE | 61,86 | 33,57 | 0,80 | 14,85 |
| MESURE | 48,89 | 30,95 | 0,82 | 12,96 |
| GEOMETRIE | 44,89 | 21,21 | 0,58 | 13,77 |

Le résultat moyen en mathématique est de 52 sur 100. La dispersion des notes attribuées aux élèves est assez importante indiquant que l'on a affaire à un groupe hétérogène par rapport à la maîtrise de la mathématique. La consistance interne de l'épreuve est acceptable (coefficient Alpha de 0,834).

Quelques items ont été très peu réussis. Il faut noter ici que certaines questions posées sont dépendantes de la réponse fournie à une autre question; il serait souhaitable d'éviter de telles situations. Les deux dernières questions ont généralement des indices de discrimination plutôt faibles.

L'éducation scientifique et technique / l'éducation sociale

Cette épreuve, notée sur 50 points, est constituée de deux champs de formation; après une analyse de l'ensemble de l'épreuve, il y aura examen des résultats obtenus par rapport à chaque champ.

Tableau 7
Résultats obtenus à l'épreuve d'éducation scientifique et technologique / d'éducation sociale

| Items | N= | Minimum | Maximum | Moyenne | Rés/100 | Corr. |
|-------|-----|---------|---------|---------|---------|-------|
| Q1A | 117 | 0,00 | 2,00 | 1,61 | 80,56 | 0,373 |
| Q1B | 117 | 0,00 | 1,00 | 0,70 | 70,09 | 0,446 |
| Q1C | 117 | 0,00 | 1,00 | 0,56 | 56,41 | 0,496 |
| Q2A | 117 | 0,00 | 0,50 | 0,21 | 41,03 | 0,212 |
| Q2B | 117 | 0,00 | 1,50 | 0,66 | 44,02 | 0,345 |
| Q2C | 117 | 0,00 | 1,00 | 0,35 | 35,04 | 0,224 |
| Q3A | 117 | 0,00 | 3,00 | 1,76 | 58,69 | 0,275 |
| Q3B1 | 117 | 0,00 | 0,50 | 0,27 | 53,85 | 0,268 |
| Q3B2 | 117 | 0,00 | 0,50 | 0,16 | 31,20 | 0,383 |
| Q4A | 117 | 0,00 | 3,00 | 1,79 | 59,83 | 0,470 |
| Q4B | 117 | 0,00 | 6,00 | 1,87 | 31,20 | 0,550 |
| Q4C | 117 | 0,00 | 1,00 | 0,47 | 47,01 | 0,462 |
| Q4D | 117 | 0,00 | 0,50 | 0,15 | 29,06 | 0,264 |
| Q5A1 | 117 | 0,00 | 0,50 | 0,33 | 66,67 | 0,202 |
| Q5A2 | 117 | 0,00 | 0,50 | 0,32 | 64,10 | 0,170 |
| Q5A3 | 117 | 0,00 | 0,50 | 0,30 | 59,83 | 0,167 |
| Q5A4 | 117 | 0,00 | 0,50 | 0,25 | 49,15 | 0,295 |
| Q5B | 117 | 0,00 | 0,50 | 0,03 | 6,84 | 0,306 |
| Q5C | 117 | 0,00 | 1,00 | 0,21 | 20,51 | 0,355 |
| Q5D | 117 | 0,00 | 2,00 | 0,44 | 22,22 | 0,385 |
| Q6A | 117 | 0,00 | 1,00 | 0,32 | 31,62 | 0,265 |
| Q6B | 117 | 0,00 | 1,00 | 0,20 | 19,66 | 0,285 |
| Q6C | 117 | 0,00 | 2,00 | 0,36 | 17,95 | 0,203 |
| Q6D | 117 | 0,00 | 2,00 | 0,24 | 11,97 | 0,277 |
| Q6E | 117 | 0,00 | 2,00 | 0,21 | 10,26 | 0,222 |
| Q7A | 117 | 0,00 | 1,00 | 0,57 | 57,26 | 0,509 |
| Q7B | 117 | 0,00 | 1,00 | 0,44 | 44,02 | 0,499 |
| Q7C | 117 | 0,00 | 1,00 | 0,38 | 38,25 | 0,356 |
| Q8 | 117 | 0,00 | 5,00 | 3,48 | 69,57 | 0,651 |
| Q9A | 117 | 0,00 | 1,00 | 0,15 | 14,53 | 0,173 |
| Q9B | 117 | 0,00 | 6,00 | 3,61 | 60,19 | 0,679 |

| Résultats | Moyenne | Ecart type | Alpha | Err-type de mesure |
|------------------|---------|------------|-------|--------------------|
| <i>Global</i> | 44,78 | 17,18 | 0,83 | 7,07 |
| <i>Habiletés</i> | | | | |
| CARACTERISATION | 42,67 | 17,20 | 0,78 | 8,14 |
| APPLICATION | 57,97 | 29,63 | 0,60 | 18,79 |
| RESOLUTION | 32,05 | 26,54 | 0,43 | 19,97 |

Contenu notionnel

| | | | | |
|---------------|-------|-------|------|-------|
| HOMME | 33,51 | 17,67 | 0,66 | 10,28 |
| NATUREL | 34,23 | 25,56 | 0,72 | 13,53 |
| VIE | 53,66 | 36,48 | 0,90 | 11,80 |
| DEVELOPPEMENT | 65,70 | 24,37 | 0,62 | 14,99 |

Trente et un éléments d'information sont recueillis par rapport à neuf questions. Le résultat moyen est plutôt faible, s'établissant à 45 sur 100. Plus de la moitié des items a un taux de réussite très faible. Plusieurs items semblent discriminer faiblement, le coefficient de corrélation item – résultat d'ensemble étant plutôt faible.

La consistance interne de l'épreuve est toutefois acceptable. Si on considère les regroupements touchant les habiletés, c'est la résolution de problèmes qui obtient le résultat le plus faible; l'habileté «application» est celle qui obtient le résultat le plus élevé.

Le tableau suivant fournit les statistiques de l'analyse d'items par rapport à chaque champ de formation inclus dans l'item.

Tableau 8
Résultats obtenus à l'épreuve d'éducation scientifique et technologique / d'éducation sociale par
champ de formation

Éducation sociale

| Items | N= | Minimum | Maximum | Moyenne | Rés/100 | Corr. |
|--------------------------|-----------|----------------|-------------------|----------------|---------------------------|--------------|
| Q1A | 117 | 0,00 | 2,00 | 1,61 | 80,56 | 0,355 |
| Q1B | 117 | 0,00 | 1,00 | 0,70 | 70,09 | 0,383 |
| Q1C | 117 | 0,00 | 1,00 | 0,56 | 56,41 | 0,476 |
| Q3A | 117 | 0,00 | 3,00 | 1,76 | 58,69 | 0,323 |
| Q3B1 | 117 | 0,00 | 0,50 | 0,27 | 53,85 | 0,305 |
| Q3B2 | 117 | 0,00 | 0,50 | 0,16 | 31,20 | 0,387 |
| Q8 | 117 | 0,00 | 5,00 | 3,48 | 69,57 | 0,797 |
| Q9A | 117 | 0,00 | 1,00 | 0,15 | 14,53 | 0,270 |
| Q9B | 117 | 0,00 | 6,00 | 3,61 | 60,19 | 0,801 |
| Résultats | | Moyenne | Ecart type | Alpha | Err-type de mesure | |
| <i>Global</i> | | 61,49 | 24,32 | 0,74 | 12,48 | |
| <i>Habilités</i> | | | | | | |
| CARACTERISATION | | 65,00 | 28,41 | 0,71 | 15,41 | |
| APPLICATION | | 57,97 | 29,63 | 0,60 | 18,79 | |
| <i>Contenu notionnel</i> | | | | | | |
| VIE | | 53,66 | 36,48 | 0,90 | 11,80 | |
| DEVELOPPEMENT | | 65,70 | 24,37 | 0,62 | 14,99 | |

Éducation scientifique et technologique

| Items | N= | Minimum | Maximum | Moyenne | Rés/100 | Corr. |
|-------|-----|---------|---------|---------|---------|-------|
| Q2A | 117 | 0,00 | 0,50 | 0,21 | 41,03 | 0,185 |
| Q2B | 117 | 0,00 | 1,50 | 0,66 | 44,02 | 0,476 |
| Q2C | 117 | 0,00 | 1,00 | 0,35 | 35,04 | 0,353 |
| Q4A | 117 | 0,00 | 3,00 | 1,79 | 59,83 | 0,512 |
| Q4B | 117 | 0,00 | 6,00 | 1,87 | 31,20 | 0,652 |
| Q4C | 117 | 0,00 | 1,00 | 0,47 | 47,01 | 0,505 |
| Q4D | 117 | 0,00 | 0,50 | 0,15 | 29,06 | 0,312 |
| Q5A1 | 117 | 0,00 | 0,50 | 0,33 | 66,67 | 0,328 |
| Q5A2 | 117 | 0,00 | 0,50 | 0,32 | 64,10 | 0,261 |
| Q5A3 | 117 | 0,00 | 0,50 | 0,30 | 59,83 | 0,274 |
| Q5A4 | 117 | 0,00 | 0,50 | 0,25 | 49,15 | 0,381 |
| Q5B | 117 | 0,00 | 0,50 | 0,03 | 6,84 | 0,349 |
| Q5C | 117 | 0,00 | 1,00 | 0,21 | 20,51 | 0,363 |
| Q5D | 117 | 0,00 | 2,00 | 0,44 | 22,22 | 0,435 |
| Q6A | 117 | 0,00 | 1,00 | 0,32 | 31,62 | 0,338 |
| Q6B | 117 | 0,00 | 1,00 | 0,20 | 19,66 | 0,351 |
| Q6C | 117 | 0,00 | 2,00 | 0,36 | 17,95 | 0,238 |
| Q6D | 117 | 0,00 | 2,00 | 0,24 | 11,97 | 0,315 |
| Q6E | 117 | 0,00 | 2,00 | 0,21 | 10,26 | 0,236 |
| Q7A | 117 | 0,00 | 1,00 | 0,57 | 57,26 | 0,418 |
| Q7B | 117 | 0,00 | 1,00 | 0,44 | 44,02 | 0,442 |
| Q7C | 117 | 0,00 | 1,00 | 0,38 | 38,25 | 0,351 |

| Résultats | Moyenne | Ecart type | Alpha | Err-type de mesure |
|--------------------------|---------|------------|-------|--------------------|
| <i>Global</i> | 33,64 | 16,63 | 0,74 | 8,46 |
| <i>Habilités</i> | | | | |
| CARACTERISATION | 33,92 | 16,81 | 0,69 | 9,30 |
| RESOLUTION | 32,05 | 26,54 | 0,43 | 19,97 |
| <i>Contenu notionnel</i> | | | | |
| HOMME | 33,51 | 17,67 | 0,66 | 10,28 |
| NATUREL | 34,23 | 25,56 | 0,72 | 13,53 |

Ainsi, si on fractionne l'épreuve en deux composantes, le nombre d'items discriminant faiblement diminue de façon significative, passant de 15 à 6. Ces résultats tendent à démontrer que l'on a bien affaire à deux épreuves différentes qui ont été intégrées pour répondre à des préoccupations d'ordre administratif.

L'éducation artistique

L'épreuve pratique d'éducation artistique a porté en juin 2000 sur le dessin. Quatre éléments d'information (il s'agit des critères de correction) sont disponibles pour réaliser l'analyse statistique.

Tableau 9
Résultats obtenus à l'épreuve d'éducation artistique

| Items | N= | Minimum | Maximum | Moyenne | Rés/100 | Corr. |
|-------|-----|---------|---------|---------|---------|-------|
| Q1 | 117 | 1,00 | 3,00 | 2,50 | 83,48 | 0,690 |
| Q2A | 117 | 2,00 | 4,00 | 3,11 | 77,78 | 0,502 |
| Q2B | 117 | 1,00 | 3,00 | 2,48 | 82,62 | 0,701 |
| Q2C | 117 | 2,00 | 5,00 | 3,37 | 67,35 | 0,790 |

| Résultats | Moyenne | Ecart type | Alpha | Err-type de mesure |
|---------------|---------|------------|-------|--------------------|
| <i>Global</i> | 76,41 | 10,88 | 0,59 | 6,94 |

Les élèves faisant partie de l'échantillon retenu ont obtenu un résultat moyen de 76 sur 100. La dispersion des élèves par rapport à ce résultat moyen est faible, indiquant que la majorité des élèves a obtenu une note se rapprochant de la moyenne générale. On doit noter ici que tous les élèves ont obtenu une note supérieure à 50 sur 100. Par ailleurs, le coefficient de cohérence interne (Alpha = 0,593) est plutôt faible; le peu d'éléments d'information compris dans l'épreuve explique en partie la valeur de cet indice. Enfin, indiquons que les coefficients indiquant la discrimination sont tous élevés, signifiant que chaque critère retenu identifie correctement les élèves en situation de maîtrise et ceux ne maîtrisant pas la tâche demandée.

L'éducation physique et sportive

L'épreuve d'éducation physique et sportive de juin 2000 comprenait les deux composantes suivantes : la vitesse (sur 60 mètres) et l'endurance (sur 2000 mètres).

Tableau 10
Résultats obtenus à l'épreuve d'éducation physique et sportive

| Items | N= | Minimum | Maximum | Moyenne | Rés/100 | Corr. |
|-------|----|---------|---------|---------|---------|-------|
| VIT_D | 94 | 1,50 | 2,50 | 2,28 | 91,06 | 0,589 |
| VIT_A | 94 | 0,50 | 1,50 | 1,37 | 91,49 | 0,607 |
| VIT_T | 94 | 2,75 | 6,00 | 4,94 | 82,27 | 0,719 |
| END_1 | 94 | 0,50 | 4,00 | 3,39 | 84,84 | 0,664 |
| END_2 | 94 | 2,25 | 6,00 | 4,86 | 81,07 | 0,580 |

| Résultats | Moyenne | Ecart type | Alpha | Err-type de mesure |
|--------------------------|---------|------------|-------|--------------------|
| <i>Global</i> | 84,22 | 10,46 | 0,51 | 7,31 |
| <i>Habiletés</i> | | | | |
| CONNAISSANCES | 88,03 | 14,97 | 0,58 | 9,71 |
| APPLICATION | 81,67 | 10,38 | 0,25 | 8,98 |
| <i>Contenu notionnel</i> | | | | |
| DEPAR | 91,06 | 16,75 | | 16,75 |
| ARRIVEE | 91,49 | 22,37 | | 22,37 |
| TEMPS | 81,67 | 10,38 | 0,25 | 8,98 |
| NATURE | 84,84 | 21,37 | | 21,37 |

Le résultat global pour cette épreuve s'établit à 84 sur 100, tous les élèves ayant obtenu une note supérieure à 50 sur 100. La valeur de l'écart type (10,4) indique que la dispersion des élèves par rapport au résultat moyen est peu importante. Le coefficient de cohérence interne de l'épreuve est moyennement faible (0,512); cependant, le peu d'éléments d'information disponible pour calculer les résultats (cinq indicateurs) justifie la valeur obtenue. Les indices de discrimination, indiquées par la corrélation item – résultat global, sont acceptables.

Comparaisons des résultats obtenus au CEP, nouveaux programmes, et au CEP, programmes intermédiaires

Les deux tableaux suivants présentent des statistiques quant à la proportion d'élèves admis au CEP pour les programmes intermédiaires (sessions de 1994 à 2000) et pour les nouveaux programmes (sessions de 1998 à 2000).

Tableau 11
Taux d'obtention du CEP – Programmes intermédiaires

| Année | Présents | % d'admis |
|-------|----------|-----------|
| 1994 | 50914 | 56,4 |
| 1995 | 58305 | 51,9 |
| 1996 | 68562 | 62,5 |
| 1997 | 74510 | 70,0 |
| 1998 | 75773 | 67,2 |
| 1999 | 80299 | 68,5 |
| 2000 | 83955 | 68,1 |

Tableau 12
Taux d'obtention du CEP – Nouveaux programmes d'études

| Année | Présents | % d'admis |
|-------|----------|-----------|
| 1998 | 996 | 92,8 |
| 1999 | 972 | 82,3 |
| 2000 | 1022 | 87,5 |

Par rapport aux programmes intermédiaires, on constate une progression constante du nombre de candidats présents à l'examen du CEP (de près de 51000 en 1994 à près de 84000 en 2000). La proportion de candidats admis semble se stabiliser à 68%. Quant aux présences des candidats au CEP, nouveaux programmes, elles se situent à environ mille. Les taux d'admission affichent plus de 80%.

On peut s'interroger sur les raisons expliquant de meilleurs succès des candidats provenant des nouveaux programmes si on les compare aux élèves des programmes intermédiaires. Est-ce que les épreuves seraient plus faciles? Est-ce que l'introduction des nouveaux programmes explique le succès des élèves? Il faudrait entreprendre des travaux de comparaison de l'instrumentation utilisée pour chaque format de CEP.

Coût et efficacité du nouveau format des épreuves

Juin 2000 marque la troisième administration du CEP, nouveaux programmes. Il y a eu, à chaque année, des activités préparatoires et des activités de suivi. C'est ainsi que des techniques pour réaliser les différentes opérations ont dû être arrêtées et maîtrisées par les nombreux partenaires impliqués : précision aux tableaux de spécification, élaboration des instruments (épreuves et grilles de correction), analyse de données recueillies auprès d'échantillons de candidats. La réglementation régissant cet examen a été révisée pour mieux tenir compte de la réalité et adapter le format à une généralisation qui arrive à grands pas; le souci des responsables demeure toutefois le respect de la rigueur et de l'équité du mécanisme d'évaluation à mettre en place pour remplir les responsabilités liées à la sanction des études primaires.

Même si les opérations prévues pour mettre sur pied l'examen 2000 du CEP, nouveaux programmes, ont été réalisées avec une maîtrise plus grande des techniques nécessaires, des améliorations aux pratiques d'élaboration de l'instrumentation ont pu être identifiées au cours des travaux réalisés à Pobè par les participants à cet atelier. La session de juin 2001 sera l'occasion d'évaluer concrètement la capacité des différents acteurs puisque le nombre de candidats sera beaucoup plus important si on compare à la situation des trois dernières sessions. Il y aura de grands défis à relever principalement par rapport à l'administration des épreuves et à leur correction.

Il semble sage de ne pas changer le format actuel du CEP, nouveaux programmes, tel que modifié après la tenue des sessions de 1998 et 1999. La session de 2001 constituera un véritable test quant au format déjà arrêté. Il sera peut-être nécessaire d'apporter des changements à la suite de cette mini généralisation.

La session 2000 du CEP, nouveaux programmes

La préparation du matériel nécessaire à la tenue de la session 2000 du CEP a été réalisée avec un souci de mieux établir des liens avec les tableaux de spécification. C'est ainsi que des items ont été élaborés par rapport aux différentes dimensions à évaluer. Sans considérer la procédure retenue comme s'appuyant sur l'existence ou le développement de banques d'items, les travaux d'élaboration de l'instrumentation s'en sont largement inspirés. C'est ainsi qu'au terme des activités réalisées en mars 2000, un matériel d'évaluation respectant les paramètres des tableaux de spécification était disponible. Toutefois, il a été impossible de réaliser immédiatement l'assemblage de ce matériel en jeux d'épreuves. C'est une autre équipe qui a dû prendre le relais. Il serait souhaitable que l'ensemble des opérations soit finalisé par l'équipe mise sur pied pour l'élaboration de l'instrumentation; un groupe réduit pourrait alors prendre la relève pour réaliser un dernier examen du matériel proposé au directeur de la DEC.

Activités complémentaires réalisées au cours de la mission

Les termes de référence développés pour cette mission de consultation ont supposé la réalisation de plusieurs tâches qui ont dû se tenir en peu de temps. Toutefois, comme cela arrive souvent dans un tel contexte, des activités complémentaires furent ajoutées pour tenir compte de la situation du moment. Nous pouvons identifier trois catégories de travaux qui se sont réalisés en parallèle à ceux exigés par les termes de référence. Les voici rapidement présentés.

Au moment de la tenue des travaux de l'atelier de Pobè, il a été nécessaire d'ajouter au programme d'activités des *sessions de formation* sur les concepts reliés à l'évaluation des apprentissages. C'est ainsi que des mini sessions de formation à différentes techniques d'analyses statistiques ont été insérées au besoin. Le consultant, à la demande expresse des participants, a préparé rapidement un lexique présentant un ensemble de concepts sur l'évaluation des apprentissages (voir l'annexe «A»).

Pour être en mesure de mettre à la disposition des participants des données statistiques qui ont apporté un éclairage complémentaire à l'analyse de l'instrumentation utilisée au CEP, nouveaux programmes, de juin dernier, il a été nécessaire d'exécuter des compilations statistiques à l'aide du logiciel utilisé depuis plus de cinq ans. La dernière version de ce programme informatique installé sur l'ordinateur utilisé à Pobè datait de plus de deux ans. Le consultant avait prévu installer une version plus récente de ce logiciel. Il a donc fallu prévoir une *mise à niveau des connaissances techniques* des personnes concernées afin de prendre en compte les modifications apportées au produit.

Au cours des deux dernières années, certains responsables du projet CLEF ont sollicité du consultant des renseignements sur la *mise en place d'un programme national d'évaluation des élèves*. Une telle activité avait eu lieu entre 1994 et 1996; elle a conduit à la réalisation d'évaluation d'échantillons de classes dans deux disciplines et la production de rapports qui ont été distribués auprès des responsables scolaires. Le sujet est encore revenu au cours de cette mission. Des rencontres ont été tenues pour préciser les activités réalisées et ce qui devrait être fait pour assurer une telle évaluation.

Activités quotidiennes

Cette section du rapport rappelle les activités réalisées au cours de la présente mission. Nous y indiquons aussi les rencontres faites dans le cadre de la réalisation des travaux convenus.

Le lundi, 22 janvier 2001

Une rencontre se tient aux bureaux du Projet CLEF à Cotonou réunissant les principaux responsables reliés aux termes de référence de la présente mission. Le pilote du Plan d'actions Évaluation des apprentissages et orientation scolaire présente les termes de références reformulés pour tenir compte de la durée qui est passée de trois semaines à deux semaines. C'est ainsi que le volet relié à la formation du personnel à l'utilisation de banques d'items pour la préparation de jeux d'épreuves a été retiré des termes de référence de la présente mission. Le consultant fait part des incertitudes qui ont été engendrées au moment de la préparation administrative de la présente mission : retard dans la prise de connaissance de la durée retenue de la consultation (réduction à deux semaines plutôt que trois communiquées auparavant, itinéraire, lieu d'hébergement, moyen de transport).

Le pilote du plan d'actions sur l'évaluation des apprentissages indique que des collègues sont réunis à Pobè depuis le milieu de la semaine précédente. On y fait l'analyse de la cohérence entre les épreuves administrées en juin dernier et les tableaux de spécification; les participants ont relevé des incohérences dans certaines épreuves. De son côté, le consultant signifie qu'il arrive aux mêmes conclusions en utilisant les bases de données qui lui ont été expédiées avant son départ pour le Bénin.

Au cours de l'après-midi, il y a rencontre au bureau du directeur de la Direction des Examens et Concours. Plusieurs points sont alors abordés. Le directeur indique que certains items faisant partie des jeux

d'épreuves proposés ont un format de présentation exigeant des ressources supplémentaires difficilement applicables dans le contexte de la généralisation du CEP, nouveaux programmes d'études. Il faut aussi avoir à l'esprit que les programmes d'études risquent d'être modifiés lorsqu'il y aura généralisation dans quatre années; les nouveaux programmes d'études font présentement l'objet d'ajustement après leur mise à l'essai. Il faut donc être prudent dans la constitution des items faisant partie d'une éventuelle banque d'items. Le directeur de la DEC souhaiterait, du moins à court terme, disposer de jeux d'épreuves mises en forme plutôt qu'avoir à choisir des items faisant partie d'une banque structurée par rapport à des tableaux de spécification. La question de l'accès à l'information doit être traitée avec soin pour s'assurer qu'il n'y a pas de fuites ou utilisation d'informations confidentielles à des fins personnelles.

Le mardi, 23 janvier 2001

Le consultant rejoint les participants à l'atelier à Pobè. Dans un premier temps, on fait un compte-rendu des travaux réalisés au cours de quatre derniers jours. Les activités ont principalement porté sur l'examen des tableaux de spécification par rapport au contenu des épreuves administrées en juin dernier. On constate que certains instruments ne respectent pas les paramètres retenus dans les tableaux de spécification. On a établi une comparaison avec les programmes d'études et il y a eu ajustement aux tableaux de spécification. Le consultant réagit aux renseignements communiqués. Le terme «connaissances» a été retenu pour désigner l'axe des domaines ou des contenus; on s'interroge sur la justesse du terme compte tenu de la signification de cet axe des tableaux de spécification; après un long débat, on retient l'expression «contenus notionnels». Par la suite, le consultant dresse la liste des habiletés contenues dans les différents tableaux de spécification et s'interroge sur la cohérence de la terminologie retenue. On décide de poursuivre la réflexion en groupes de travail intégrant des membres de tous les champs de formation.

Le consultant se joint à l'équipe technique pour réaliser les compilations statistiques à partir des fichiers de données saisis par le personnel du Ministère. Un examen des bases de données doit tout d'abord être réalisé. Par la suite, des analyses d'items et des résultats moyens sont produits. Ce travail permet au personnel technique rattaché au projet de réviser les concepts reliés à une telle opération de compilations statistiques.

Le mercredi, 24 janvier 2001

Au cours de l'avant-midi, l'équipe technique termine les compilations statistiques et fait l'impression de rapports qui seront distribués aux participants au moment de l'analyse statistique des épreuves du CEP de juin 2000. Les participants à l'atelier, de leur côté, poursuivent la réflexion sur les tableaux de spécification, en particulier sur les habiletés à retenir.

Par la suite, il y a échanges en plénière. À la fin de la journée, le consultant remet copies des rapports statistiques et présente un sommaire des résultats obtenus.

Le jeudi, 25 janvier 2001

L'avant-midi est consacré à examiner les rapports statistiques et à présenter des explications quant au sens à donner aux différents indices présents sur les rapports rendus disponibles aux participants. Ceux-ci souhaitent avoir plus d'information sur les concepts statistiques appliqués à l'évaluation des apprentissages. On désirerait aussi approfondir les connaissances des techniques pouvant être utilisées pour mettre en

place des banques d'items dans le but de préparer des épreuves pour le CEP. Le consultant ne peut qu'indiquer que ces préoccupations dépassent malheureusement les termes de référence de la présente mission de consultation.

À la suite des explications fournies par le consultant sur les indices statistiques, les équipes se mettent à l'examen des résultats obtenus à chaque épreuve administrée en juin 2000.

Le vendredi, 26 janvier 2001

Les travaux en ateliers se poursuivent. Le consultant fournit des renseignements supplémentaires sur le sens à donner à certains indices statistiques apparaissant sur les rapports produits au début de la tenue de la mission. À la suite de demandes pressantes des participants d'obtenir des définitions de concepts reliés à l'évaluation des apprentissages, le consultant prépare un document reprenant des renseignements contenus dans l'aide en contexte du logiciel EduStat (ce document est reproduit à l'annexe «A»).

Une rencontre est aussi organisée avec le personnel de la DEC qui participe à l'atelier. Dans un premier temps, nous avons abordé la question des statistiques produites par cette Direction à la suite de l'administration du CEP de juin 2000. Le consultant pourra examiner ces statistiques la semaine prochaine; on rendra aussi disponible les copies d'épreuves utilisées pour le CEP, programmes intermédiaires. Par la suite, nous avons abordé les prochaines étapes à réaliser. La session de juin 2001 accueillera beaucoup plus de candidats au CEP, nouveaux programmes (environ 6000). On doit préparer des jeux d'épreuves. On relève qu'il y aura plusieurs correcteurs nouveaux qui devront se joindre pour réaliser les opérations nécessaires; il faudra assurer leur formation. Le troisième point touché vise les coûts reliés à l'administration des épreuves. Le format des épreuves doit respecter l'esprit des nouveaux programmes; on convient qu'il peut y avoir des coûts plus importants que ceux déboursés pour l'administration du CEP, programmes intermédiaires. Il faut tenter de trouver des formules qui, tout en garantissant la rigueur de l'évaluation et le respect de l'esprit des nouveaux programmes, nécessiteront des ressources qu'il sera possible d'assumer. Dans cet ordre d'idées, on entrevoit, par exemple, de préparer un guide d'administration intégré, de préparer du matériel qui pourra être utilisé d'une année à l'autre.

À la fin de la journée, le consultant se rend, à la demande du bureau CLEF de Porto-Novo, à une rencontre avec le personnel du CLEF de Cotonou (messieurs Hébert et Camara) ainsi que madame Alima Boukary-Marcos, du bureau de Porto-Novo. On y discute d'une possible étude à réaliser concernant un mécanisme d'évaluation des apprentissages qui devait être déjà réalisé; le consultant avait fourni l'an dernier des avis au personnel du CLEF par rapport à ce projet. On convient d'examiner la question et de proposer des termes de référence en conséquence.

Le samedi, 27 janvier 2001

L'atelier se poursuit à Pobè en avant-midi. Le consultant revient avec les membres de certaines équipes sur la signification à accorder à différents indices statistiques. Il y a aussi une rencontre avec le Pilote du plan d'actions pour discuter des suites à donner à la rencontre tenue la veille à Porto-Novo.

Au cours de l'après-midi, il y a rencontre à Porto-Novo avec madame Alima Boukary-Marcos sur la rédaction de termes de référence pour assurer une suite aux préoccupations du CLEF concernant le volet

de programme national d'évaluation des élèves. Le consultant propose des éléments d'information reliés aux tâches à réaliser et aux résultats à produire. De son côté, le Pilote indique les opérations à prévoir et les ressources à mettre en place.

Le lundi, 29 janvier 2001

Au cours de l'avant-midi, il y a eu suite des travaux en atelier pour l'examen des résultats obtenus par les élèves du CEP, nouveaux programmes. Des questions ont été posées au consultant concernant le sens à donner à l'écart type et la façon d'obtenir cette statistique. À l'aide d'un exemple simple, il a été possible d'explicitier les notions d'écart type, de résultat moyen, de médiane, de mode et d'erreur type d'échantillonnage.

En après-midi, le groupe responsable de l'analyse des résultats obtenus à l'épreuve de mathématique fait son rapport. Cette présentation donne lieu à plusieurs questions de clarification. Comme la journée est déjà écoulée, il est convenu de remettre au lendemain l'examen des autres épreuves. Le consultant quitte Pobè à 19 :30 heures, la voiture mise à sa disposition par le projet CLEF ayant accusé un retard important...

Le mardi, 30 janvier 2001

La journée a été consacrée au compte-rendu des travaux réalisés par les équipes par rapport aux épreuves suivantes : l'éducation physique et sportive, l'éducation artistique, la lecture, l'expression écrite et l'orthographe. Les débats qui ont suivis la présentation des résultats des analyses ont été très animés. Il faudra que les équipes complètent leur rapport en tenant compte des remarques formulées par les participants.

Le mercredi, 31 janvier 2001

L'avant-midi du consultant a été consacré à examiner avec l'aide technique du CLEF à Porto-Novo les termes de référence d'une activité d'évaluation qui devrait se dérouler d'ici juillet 2001. Par la suite, le consultant a rejoint les collaborateurs réunis à Pobè. Il y a eu présentation des travaux réalisés par l'équipe responsable de l'examen de l'instrumentation et des résultats obtenus à l'épreuve touchant les champs de formation Éducation sociale et Éducation scientifique et technologique.

Le jeudi, 1^{er} février 2001

Au cours de l'avant-midi, le consultant a travaillé avec les membres de l'équipe de français sur la grille de correction de l'épreuve d'expression écrite. Par la suite, il s'est mis à la disposition de l'équipe technique pour examiner le fonctionnement du logiciel EduStat par rapport à la fabrication de graphiques. À la reprise des travaux en après-midi, il y a eu plénière au cours de laquelle l'équipe du champ de formation de l'Éducation artistique a présenté une grille de correction pour un épreuve de dessin. Par la suite, le consultant a poursuivi le travail avec l'équipe technique dans le but de rendre le personnel plus autonome dans l'utilisation du logiciel utilisé pour réaliser les analyses statistiques.

Le vendredi, 2 février 2001

Il y a rencontre d'évaluation de la mission du consultant au cours de l'avant-midi; cette réunion se tient au bureau du CLEF à Porto-Novo. Un compte-rendu des travaux réalisés a pu être fait à cette occasion. Le consultant a indiqué son incompréhension des problèmes administratifs qui se sont manifestés au cours de ce séjour. À la requête du représentant de l'USAID, des exemples de «ratés» administratifs furent fournis. On assure le consultant que cela ne se reproduira plus.

Après cette rencontre d'évaluation, le consultant se rend à Pobè pour participer à la session de clôture de l'atelier. Par la suite, il y a rencontre avec les membres retenus pour conduire les activités devant être réalisées dans le cadre d'une deuxième évaluation de classes de CP et de CM1 en français et en mathématique.

Suite des opérations

Les travaux réalisés par les participants à l'atelier de Pobè avant l'arrivée du consultant et au cours de son séjour ont permis d'une part de dresser un bilan de la session de juin 2000 du CEP et d'autre part de préparer les activités à prévoir pour la tenue du prochain CEP, nouveaux programmes. Nous présentons ci-dessous un aperçu des opérations à réaliser et des suites qu'il est possible d'entrevoir.

- *Expertise nationale en techniques d'analyses du rendement des élèves.* Comme nous l'indiquions plus haut, il s'agit de la troisième édition de cette forme d'analyse du matériel et des procédures retenus pour la tenue du CEP, nouveaux programmes. La réalisation de ces examens successifs s'appuie sur des données recueillies auprès d'échantillons de candidats. Les participants aux différents ateliers organisés depuis 1999 ont pu acquérir des connaissances techniques reliées à l'analyse statistique adaptée à l'évaluation des apprentissages. Un équipe technique a développé l'expertise nécessaire à la réalisation des travaux exigés pour disposer de l'information requise. Ils sont de plus en plus nombreux les participants à maîtriser les concepts et les techniques permettant une analyse rigoureuse des résultats obtenus à des épreuves administrées dans le cadre du CEP.
- *Améliorations à apporter.* Même si on en est au troisième cycle d'analyse du CEP, nouveaux programmes, on constate qu'il y a encore des lacunes à corriger au matériel et aux procédures pour rendre cet examen plus rigoureux. Les analyses réalisées au cours du présent atelier a permis d'identifier des aspects à corriger pour disposer d'une instrumentation améliorée. Chaque équipe de travail a rédigé des rapports rendant compte des observations faites par les participants. Il faudra s'y référer au moment de l'élaboration des jeux d'épreuves nécessaires à la tenue du CEP de juin 2001.
- *Uniformisation de l'instrumentation.* Les ateliers qui se sont réalisés en 1999 et en 2000 ont mis l'accent sur les travaux en équipes formées de spécialistes affectés à l'analyse d'un seul champ de formation. Ce mode de fonctionnement a permis un examen en profondeur de chaque champ de formation. On a toutefois constaté que les travaux ont évolué à des rythmes différents compte tenu de la composition des équipes et des particularités des champs de formation. Il est apparu important de réaliser des travaux regroupant des participants provenant des différents champs de formation. Il faudra tenter d'uniformiser certains aspects de l'instrumentation tout en étant conscient qu'il faut respecter les particularités des champs de formation.

- *Préparation de l'instrumentation 2001.* Il faudrait songer rapidement à la mise en place des ressources nécessaires à la préparation du matériel nécessaire à la tenue du CEP 2001. Il serait souhaitable de réserver assez de temps pour permettre à une équipe réduite de réaliser une révision finale de l'instrumentation en s'efforçant tout particulièrement d'uniformiser la présentation des documents. On pourrait tenter d'intégrer les différents guides d'administration dans le but d'éviter la redondance dans la diffusion de l'information.
- *Prochaines compilations touchant le CEP 2001.* À la suite de l'administration du CEP 2001, il semble nécessaire de prévoir une opération de recorection d'un échantillon de candidats, d'une entrée de ces données et de la réalisation de compilations statistiques. Nous l'avons déjà noté; il y aura beaucoup plus de candidats au CEP 2001, nouveaux programmes, en provenance d'un nombre plus important d'écoles. Nous proposons donc de retenir un échantillon plus important pour réaliser cette analyse; le nombre de candidats à retenir devrait être fixé à 300. Ceci permettra de réaliser des compilations plus fiables.
- *Disponibilité des ressources informatiques à la DEC.* Au cours de cet atelier, le personnel de la DEC a participé activement aux travaux d'analyse des instruments utilisés au CEP. D'un autre côté, une demande d'acquisition de matériel informatique a été formulée par cette Direction. Si on veut être en mesure de généraliser à moyen terme le format du CEP, nouveaux programmes, il sera important que la DEC dispose des outils nécessaires, à savoir l'équipement, les logiciels et les compétences techniques. Il faudrait examiner cette question dans un avenir très proche.
- *Ressources informatiques mises à la disposition du plan d'actions «Évaluation».* Le plan d'actions portant sur l'évaluation se voit confier de nombreux travaux techniques qui exigent le recours à des outils informatiques permettant l'exploitation des technologies supposant la disponibilité d'appareils performants. Les ordinateurs mis à la disposition du personnel de ce plan d'actions permettent de réaliser les travaux, il est vrai; mais, c'est souvent au prix de nombreuses difficultés. Il faudrait songer à renouveler certains appareils et y adjoindre des périphériques adaptés à ce champ d'expertise (par exemple, «scanner», graveur de CD, modem, unité de sauvegarde de type ZIP). Par ailleurs, des échanges d'information via le courrier électronique (par exemple, expédition de fichiers de données, de bases de données entre le pilote et le consultant) permettrait de diminuer les coûts puisque l'on pourrait réduire la durée pour de la consultation sur place.

Annexe « A » Quelques définitions de concepts reliés à l'évaluation⁶

Une **banque d'item** est un fichier contenant un ensemble d'items validés et dont on connaît les caractéristiques docimologiques.

Le **centile** est une mesure de position utilisée pour diviser une série de données ordonnées en cent parties égales à l'aide de quatre-vingt-dix-neuf valeurs.

Une **clé de correction** est l'ensemble des réponses correctes ou admises pour les items d'un instrument de mesure.

Par **codification**, on entend l'opération qui consiste à attribuer des codes à un élément d'information dans le but d'être en mesure d'en faire la compilation.

Le **coefficient Alpha** permet de mesurer la fidélité d'un test ou d'une épreuve.

Le **coefficient de variation** indique le degré d'homogénéité des données d'une série, c'est-à-dire si les données sont semblables et se regroupent autour de la moyenne. La formule de calcul fait intervenir l'écart type et la moyenne.

La **cohérence interne** est le degré avec lequel les items ou les parties d'un instrument mesurent effectivement la même caractéristique ou le même objectif.

La **consistance interne** est le degré avec lequel les items ou les parties d'un instrument mesurent effectivement la même caractéristique ou le même objectif.

Une **corrélation** est le rapport entre deux ou quelques phénomènes qui varient en relation l'un avec l'autre.

La **cote z** mesure la distance entre une donnée de la série et la moyenne de cette série mesurée en écart-type. On obtient cette statistique en faisant la différence entre la moyenne analysée et la moyenne globale, le tout divisé par l'écart-type.

⁶ Information tirée de l'aide en contexte du logiciel EduStat (EduStat © est une marque de commerce enregistré pour EDUCAN Inc.)

Un **critère** est le principe auquel on se réfère pour porter un jugement, une appréciation. Caractéristique d'une performance ou d'un produit.

On peut dire que l'on a une **différence significative** lorsque la différence entre les résultats obtenus auprès de deux échantillons, ou entre les résultats d'un échantillon et ceux de la population dont il est extrait, ne peut pas être attribuée au hasard.

La **dispersion** est le degré de variation ou d'étalement des valeurs observées d'une variation autour d'une valeur de tendance centrale.

Le terme **distribution** désigne généralement la série des valeurs ou données correspondant à une variable quantitative.

La **docimologie** est la science de la mesure et de l'évaluation qui a comme objet d'étude l'élaboration d'instruments de mesure, l'interprétation des informations obtenues par la mesure, le comportement des examinés, de même que l'élaboration et la réalisation de plans d'évaluation.

L'**écart type** est la mesure de dispersion la plus usuelle. Cette caractéristique mesure la concentration de la distribution d'une variable autour de sa moyenne. Techniquement, l'écart type est la racine carrée positive de la variance.

Un **échantillon** doit permettre de reproduire le plus fidèlement possible les caractéristiques de la population étudiée. Dans un tel cas, on dit que l'échantillon est représentatif d'une population, c'est-à-dire qu'il constitue un portrait fidèle de celle-ci. Les individus ou les éléments de l'échantillon se retrouvent donc, par rapport à un certain nombre de caractéristiques ou de variables, dans les mêmes proportions que dans la population elle-même. Lorsque le responsable d'une recherche a recours à un échantillon plutôt qu'à tous les individus de la population étudiée, il doit accepter le fait que les résultats obtenus comportent une certaine marge d'erreur qui provient du fait qu'il analyse seulement la partie de la population incluse dans l'échantillon. Si on doit généraliser les résultats à la population totale, on peut estimer cette erreur que si on met en place une procédure d'échantillonnage probabiliste, c'est-à-dire que les individus retenus dans l'échantillon sont choisis au hasard.

L'**échelle de mesure** est la façon de mesurer une variable en vue d'obtenir des données. Il y a plusieurs types d'échelles : nominale, ordinale, d'intervalles, de rapport.

L'**échelle de rapport** est une échelle de mesure d'une variable quantitative dans laquelle le zéro est absolu. Une telle échelle permet de comparer les diverses mesures et de trouver des rapports entre elles.

L'**échelle d'intervalles** est une échelle de mesure d'une variable quantitative dans laquelle le zéro est relatif. Ainsi, une telle échelle permet de mesurer la distance entre n'importe quelles deux mesures.

L'**échelle nominale** est une échelle de mesure d'une variable qualitative (par exemple, le sexe). En d'autres termes, l'échelle nominale permet le classement d'individus en catégories de sorte que chaque individu appartienne à une seule catégorie.

L'**échelle ordinale** est une échelle de mesure d'une variable qualitative dont les modalités représentent un ordre (par exemple, l'appréciation d'un spectacle). Une telle échelle permet donc d'établir un ordre linéaire entre les données recueillies.

L'**effet de Halo** est l'influence qu'exercent divers facteurs d'ordre affectif sur l'objectivité de la correction.

Une **épreuve** est un instrument de mesure composé d'un ensemble d'items choisis en fonction des objectifs dont on veut mesurer l'atteinte.

L'**erreur type de mesure** est la différence entre une valeur observée et la valeur « vraie » ou réelle de ce que l'on mesure, due au manque de fidélité du test ou du processus de mesure

L'**erreur type d'échantillonnage** est une statistique permettant d'estimer la marge d'erreur qu'il faut accepter si l'on souhaite généraliser à la population ce qui a été observé dans l'échantillon prélevé. Cette erreur est estimable s'il s'agit d'un échantillon probabiliste.

L'**étendue** est l'écart entre la plus grande valeur et la plus petite valeur d'une série.

L'**évaluation** est une démarche qui consiste à porter un jugement sur un programme ou sur le rendement d'un élève ou d'un groupe d'élèves à partir d'informations en vue d'une décision.

La **fidélité** est la qualité d'un instrument qui permet théoriquement de recueillir la même mesure d'un phénomène ou d'un objet, à deux moments différents, c'est-à-dire de produire des résultats constants.

La **fréquence** ou l'**effectif** d'une catégorie est le nombre d'individus ou d'éléments pour lesquels les données correspondent à cette catégorie.

L'**indépendance locale** est l'absence de liens entre des items faisant partie d'une épreuve ne rendant pas nécessaire le succès à un ou plusieurs items pour pouvoir répondre correctement à un autre item.

Un **indicateur** est une mesure non pas du concept étudié lui-même, mais de l'une de ses manifestations.

L'**indice de difficulté** est un nombre indiquant la proportion des élèves ayant réussi un item par rapport à l'ensemble des élèves qui ont répondu à cet item.

L'**indice de discrimination** est un nombre exprimant le pouvoir qu'a un item ou un test de différencier les élèves selon qu'il possèdent plus ou moins la caractéristique mesurée.

L'**inférence statistique** est la méthode par laquelle des observations faites sur un petit nombre d'individus sont généralisées à toute une population.

Un **item** est l'élément d'un instrument de mesure, généralement présenté sous la forme d'un problème à résoudre, d'une question à répondre ou d'une tâche plus ou moins complexe à exécuter.

La **médiane** est une mesure de tendance centrale qui divise une suite ordonnée de scores en deux sous-ensembles égaux.

Le **mode** est le nombre ou le code qui apparaît le plus fréquemment dans une série de scores ou de codes.

La **moyenne** est le nombre indiquant la tendance centrale d'un ensemble de données numériques et dont la grandeur dépend de chacun des éléments de l'ensemble.

La **notation** est l'action d'attribuer une note (chiffre ou %) ou une cote (lettre ou symbole) ou encore de donner une appréciation ou un commentaire sur les apprentissages scolaires ou le développement général de l'élève.

Une **question d'opinion** est l'information demandée à un individu à l'aide de réponses qui se situent sur une échelle d'intervalles pour préciser quantitativement les différences qui séparent les entités mesurées.

Le **rang centile** est une mesure de position qui situe une valeur en fonction du nombre de valeurs (exprimé en pourcentage) qui lui sont inférieures.

La **représentativité** est la qualité d'un échantillon composé de façon à contenir les mêmes caractéristiques que la population dont il est extrait; ceci permet alors d'extrapoler les observations à la population.

La **saisie** est l'opération qui consiste à entrer à l'ordinateur des informations recueillies à l'aide d'une épreuve ou d'un questionnaire de façon à les organiser selon un format préterminé.

Un **sondage** est la consultation d'une partie de la population.

La **statistique** est la science du traitement de données et de leur interprétation. La statistique descriptive constitue l'ensemble des procédures et des techniques qui ont pour but de décrire un phénomène observé. Les statistiques est le terme qui désignent ordinairement des données brutes ou partiellement traitées, généralement publiées sous forme de tableaux, sans analyse.

Les **strates** sont des regroupements homogènes d'individus ou d'éléments d'une population selon un critère qui permet une classification exhaustive des individus ou des éléments et des catégories mutuellement exclusives.

Un **tableau de spécification** est une présentation ordonnée de l'ensemble des notions visées par un instrument de mesure ou un programme, qui inclut des indications sur le niveau taxonomique de l'apprentissage relié à chaque notion et sur l'importance relative d'une notion ou d'un sous-ensemble de notions par rapport à l'ensemble total.

Un **tableur** est un programme de création et de manipulation interactives de tableaux numériques visualisés.

Par **taxonomie**, on entend une classification hiérarchique des comportements cognitifs, affectifs ou moteurs, établie sur la base d'un ou de plusieurs critères.

Un **test** est un instrument de mesure dont l'expérimentation a démontré la fidélité, l'objectivité et la validité et dont l'utilisation est standardisée.

Un **test statistique** est une procédure visant à déterminer si des observations faites sur un échantillon sont valables pour toute la population et s'il existe une relation entre deux variables.

Le **test t** est un test statistique utilisé pour mesurer si la différence observée entre les moyennes de deux séries est significative.

La **validité** est la qualité d'un instrument qui mesure bien ce qu'il prétend mesurer, rien de plus.

Une **variable** est une caractéristique de personnes, d'objets ou de situations liée à un concept et pouvant prendre diverses valeurs.

Une **variable continue** est une variable quantitative dont la donnée numérique est le résultat d'une mesure relative à un étalon, sa valeur étant alors un nombre réel.

Une **variable discrète** est une variable quantitative dont la donnée numérique est le résultat d'un décompte en entier.

Une **variable quantitative** est celle dont les données qui s'y rapportent sont le résultat d'un décompte ou d'une mesure numérique, faisant appel aux propriétés arithmétique des nombres.

La **variance** est la moyenne des carrés des écarts entre chaque donnée et la moyenne.

La **vérification des données** est une opération permettant d'évaluer les données recueillies pour s'assurer qu'elles sont utilisables pour l'analyse.

Analyse d'items

La procédure Analyse d'items concerne les réponses à choix multiple ou sur une échelle continue (variables de type crédit partiel). Ce type d'analyse permet l'estimation de la qualité métrique de chaque item faisant partie d'un test ou d'une épreuve. Pour chaque item d'une épreuve à choix multiple, le rapport présente le nombre d'individus, la proportion de ceux qui ont réussi l'item, le coefficient de corrélation item/test (corrélacion bisériale ou du point bisérial), l'écart type. Il est possible de choisir de retenir une formule statistique dite « corrigée », c'est-à-dire le retrait des valeurs de l'item analysé au moment de l'établissement des statistiques. Il est aussi possible de faire afficher, pour les items à choix multiple, les pourcentages de réussite à l'ensemble du test des individus qui ont réussi l'item et de ceux qui l'ont échoué. Pour chaque item d'une épreuve à échelle continue, le rapport fournit le nombre d'individus, les minimum et maximum observés, la moyenne obtenue, l'expression de cette moyenne en pourcentage ainsi que le coefficient de corrélation item/test (corrélacion de Pearson). Il est aussi possible de faire préparer le rapport en présentant les items par rapport aux regroupements prévus dans le fichier descripteur. Pour l'ensemble de l'épreuve et par regroupement d'items, il y a calcul le coefficient de consistance interne Alpha de Cronbach, l'écart type et l'erreur type de mesure.

Corrélacion bisériale

La corrélation bisériale indique le lien qui existe entre la réponse fournie à un item et le résultat obtenu à l'épreuve. Pour réaliser ce calcul, il y a constitution, pour chaque item, de deux groupes d'individus : ceux qui ont réussi l'item et ceux qui l'ont échoué. Le programme compare alors les résultats obtenus à l'épreuve par ces deux groupes de personnes. Un taux de corrélation bisériale positivement élevé indique habituellement que ce sont les personnes les plus performantes qui ont réussi l'item et les moins performantes qui l'ont échoué. Il faut néanmoins considérer la difficulté de l'item : si celui-ci a été réussi ou échoué par la presque totalité des individus, le taux de corrélation peut alors prendre des valeurs extrêmes qu'il est difficile d'interpréter.

Les formules utilisées pour le calcul de la corrélation du point-bisérial et de la corrélation bisériale diffèrent. Aussi, les coefficients obtenus ne sont pas les mêmes pour le même item d'une même épreuve. Ordinairement, la valeur limite acceptable pour un coefficient bisérial est de 0,30 alors que celle pour un coefficient du point-bisérial est de 0,20.

Il est possible de retenir une formule statistique dite « corrigée », c'est-à-dire le retrait des valeurs de l'item analysé au moment de l'établissement des statistiques.

Corrélation de Pearson

Nous pouvons affirmer qu'il existe une relation entre deux ou plusieurs variables si leurs variations sont reliées l'une à l'autre. Le coefficient de corrélation indique donc le degré d'association existant entre deux variables. Une façon d'interpréter un coefficient de corrélation calculé à partir de deux séries de données est d'élever le coefficient obtenu au carré; le résultat de cette opération indiquera la proportion de la variation dans une variable qui s'explique par la présence de l'autre variable (on parle alors de coefficient de détermination). Ainsi, si on a obtenu une corrélation de 0,8 entre deux séries de données, on peut croire que 64% ($0,8 * 0,8$) de la variation dans une série de données s'explique par la présence de l'autre série.

L'écart type

L'écart type est la mesure de dispersion la plus usuelle. Cette caractéristique mesure la concentration de la distribution d'une variable autour de sa moyenne. Techniquement, l'écart type est la racine carrée positive de la variance.

La variance

La variance est la moyenne des carrés des écarts entre chaque donnée et la moyenne.

L'erreur type de mesure

C'est la différence entre une valeur observée et la valeur « vraie » ou réelle de ce que l'on mesure, due au manque de fidélité du test ou du processus de mesure. Cette statistique peut être fournie pour l'ensemble d'un groupe d'individu. Le calcul de cette statistique tient compte de l'écart type et de la valeur du coefficient Alpha; voici la formule utilisée :

L'erreur type de mesure = l'écart type * (la racine carrée de $1.000 -$ la valeur de l'Alpha)

Le coefficient Alpha de Cronbach

La qualité qu'a un instrument de mesurer avec la même exactitude à chaque administration réalisée dans des conditions semblables se nomme la «fidélité». Il existe plusieurs méthodes pour apprécier la fidélité d'un instrument. C'est ainsi que le calcul d'un coefficient de consistance interne est souvent retenu. Ce coefficient a pour but de refléter le degré d'association ou de corrélation qui existe entre les variables. Plusieurs coefficients peuvent être produits pour rendre compte de la consistance interne. La méthode Kuder-Richardson (KR-20) et la méthode Alpha de Cronbach sont les plus courantes. Au moment de l'analyse d'items, le logiciel utilisé produit le coefficient Alpha qui convient aux épreuves contenant des items à choix multiple ou sur une échelle continue.

Janvier 2001

Annexe « B » Tableaux de spécification révisés

CHAMP DE FORMATION MATHÉMATIQUE
TABLEAU DE SPECIFICATION CEP (Nouvelle version)

| HABILETES CONTENUS DISCIPLINAIRES | STRUCTURATION 40% | MATHÉMATISATION 20% | OPERATION 25% | SYNTHÈSE 15% | PONDERATION 100% |
|--|------------------------------|--------------------------------|--------------------------|-------------------------|-----------------------------|
| ARITHMÉTIQUE 45% | 1 18% 11 points | 2 9% 5 points | 3 11% 7 points | 4 7% 4 points | 45% 27 points |
| MESURE 25% | 5 10% 6 points | 6 5% 3 points | 7 6% 4 points | 8 4% 2 points | 25% 15 points |
| GÉOMÉTRIE 30% | 9 12% 7 points | 10 6% 4 points | 11 8% 5 points | 12 4% 2 points | 30% 18 points |
| PONDERATION 100% | 40% 24 points | 20% 12 points | 25% 16 points | 15% 8 points | 100% 60 points |

NB : Le présent tableau de spécification a été bâti à partir du programme actuel de mathématique de CM2
 POBE, le 19/01/01

CHAMP DE FORMATION ES / EST

TABLEAU DE SPECIFICATION CEP (Nouvelle version)

| HABILETES CONTENUS DISCIPLINAIRES | CARACTERISATION | MANIPULATION | APPLICATION | RESOLUTION DE PROBLEMES | PONDERATION |
|--|-----------------|-----------------|----------------|----------------------------|--------------|
| HOMME | 1 09 pts 18% | 2 | 3 | 4 1,5 pts 3% | 10,5 pts 21% |
| ENVIRONNEMENT NATUREL | 5 4,5 pts 9% | 6 | 7 | 8 1,5 pts 3% | 6 pts 12% |
| ENVIRONNEMENT TECHNOLOGIE | 9 6 pts 12% | 10 6 pts 12% | 11 | 12 1,5 pts 3% | 13,5 pts 27% |
| VIE SOCIALE | 13 3 pts 6% | 14 | 15 2 pts 4% | 16 2 pts 4% | 7 pts 14% |
| SOCIETE DEVELOPPEMENT ET ENVIRONNEMENT | 17 5 pts 10% | 18 | 19 4 pts 8% | 20 4 pts 8% | 13 pts 26% |
| LANGUE ET CULTURE | 21 | 22 | 23 | 24 | |
| PONDERATION | 27,5 pts 55% | 6 pts 12% | 6 pts 12% | 10,5 pts 21% | 50 pts 100% |

CHAMP DE FORMATION EPS
TABLEAU DE SPECIFICATION CEP (Nouvelle version)

| Contenu \ HABILITES | CONNAISSANCES | | APPLICATION | | PONDERATION | |
|---------------------|---------------|-----|-------------|-----|-------------|------|
| | | | | | | |
| VITESSE | 4 pts | 40% | 6 pts | 60% | 10 pts | 100% |
| GRIMPER | 4 pts | 40% | 6 pts | 60% | 10 pts | 100% |
| LANCER | 4 pts | 40% | 6 pts | 60% | 10 pts | 100% |
| HAUTEUR | 4 pts | 40% | 6 pts | 60% | 10 pts | 100% |
| ENCHAINEMENT | 4 pts | 40% | 6 pts | 60% | 10 pts | 100% |
| ENDURANCE | 4 pts | 40% | 6 pts | 60% | 10 pts | 100% |

CHAMP DE FORMATION EA
TABLEAU DE SPECIFICATION CEP (Nouvelle version)

| HABILETES | COMPREHENSION | | REALISATION OU EXECUTION | | JUGEMENT | | PONDERATION | |
|---------------------------------|----------------------|-------|---------------------------------|--------|-----------------|---|--------------------|----|
| CONTENUS DISCIPLINAIRES | | | | | | | | |
| EPREUVE PRATIQUE : | | | | | | | | |
| DESSIN OU COUTURE | 20% | 3 pts | 80% | 12 pts | | | 100% pts | 15 |
| EPREUVE ORALE A OPTION : | | | | | | | | |
| CHANT OU CONTE | 20% | 4 pts | 75% | 15 pts | 5% pt | 1 | 100% pts | 20 |

CHAMP DE FORMATION FRANCAIS
TABLEAU DE SPECIFICATION CEP (Nouvelle version)

| HABILETES EPREUVES | COMPREHENSION | PRODUCTION | MAITRISE DU CODE GRAMMATICAL ET/OU ORTHOGRAPHIQUE | PONDERATION |
|--|----------------------|-------------------|--|--------------------|
| COMMUNICATION ORALE/POESIE | 10 pts 12% | 6 pts 7% | 4 pts 5% | 20 pts 24% |
| LECTURE | 30 pts 35% | | | 30 pts 35% |
| EXPRESSION ECRITE/ORTHOGRAPHE | | 25 pts 29% | 10 pts 12% | 35 pts 41% |
| PONDERATION | 40 pts 47% | 31 pts 36% | 14 pts 17% | 85 pts 100% |

A – COMMUNICATION ORALE

| HABILETES CONTENUS NOTIONNELS | INFORMATION | ARGUMENTATION | REACTION A ... | PRODUCTION ORALE | MAITRISE DU CODE GRAM- MATICAL OU ORTOGRAPHIQUE | PONDERATION |
|--|--------------------|----------------------|---------------------------|-----------------------------|--|--------------------|
| VOCABULAIRE | 3 pts 15% | 4 pts 20% | 3 pts 15% | | | 10 pts 50% |
| POESIE/ EXPRESSION ORALE | | | 3 pts 15% | 3pts 15% | | 6 pts 30% |
| GRAMMAIRE/ CONJUGAISON | | | | | 4 pts 20% | 4 pts 20% |
| PONDERATION | 3 pts 15% | 4 pts 20% | 6 pts 30% | 3 pts 15% | 4 pts 20% | 20 pts 100% |

B – LECTURE

| HABILETES CONTENUS NOTIONNELS | SELECTION D'INFORMATIONS | | CONSTRUCTION DE SENS | | INFERENCE | | MAITRISE DU CODE GRAMMATICAL ET/OU ORTHOGRAPHIQUE | | PONDERATION | |
|--|---------------------------------|-----|-----------------------------|-----|------------------|-----|--|-----|--------------------|-----|
| VOCABULAIRE | 10 pts | 32% | 12 40% | pts | 4 pts | 14% | | | 26 86% | pts |
| GRAMMAIRE ET/OU CONJUGAISON | | | | | | | 4 pts | 14% | 4 14% | pts |
| PONDERATION | 10 pts | 32% | 12 40% | pts | 4 pts | 14% | 4 pts | 14% | 30 100% | pts |

C – ORTHOGRAPHE

| HABILETES CONTENUS NOTIONNELS | MAITRISE DU CODE ORTHOGRAPHIQUE | MAITRISE DU CODE GRAMMATICAL | PONDERATION |
|--|--|---|--------------------|
| ORTHOGRAPHE GRAMMATICALE | | 7 pts 70% | 7pts 70% |
| ORTHOGRAPHE D'USAGE | 3 pts 30% | | 3pts 30% |
| PONDERATION | 3 pts 30% | 7 pts 70% | 10pts 100% |

Annexe « C » Rapport sommaire

Projet CLEF – Bénin Rapport sommaire de mission de Léo Laroche (22-01-01 au 02-02-01)

L'examen du CEP 2000 : analyse de l'instrumentation utilisée et des résultats obtenus

Ce rapport sommaire établit un résumé des constats faits par le consultant dans le cadre de la présente mission. Le rapport de mission contiendra des renseignements plus complets sur chacun des points abordés⁷

Réalisation d'activités préparatoires

Le consultant a reçu, avant son départ pour le Bénin, une série de fichiers contenant des échantillons de notations à sept épreuves⁸ administrées en juin 2000 dans le cadre du CEP, nouveaux programmes d'études. Les bases de données ont été examinées; il y a eu production de statistiques pour pouvoir se prononcer sur les qualités docimologiques de ces instruments.

Cohérence entre la définition de domaines et les tableaux de spécification

Avant l'arrivée du consultant au Bénin, il y a eu des activités réalisées par les responsables béninois touchant l'examen de l'instrumentation utilisée en juin 2000 par rapport aux différents tableaux de spécification élaborés en mars 2000. On y a constaté que certaines épreuves ne respectent pas complètement les pondérations comprises dans ces tableaux. Il faut noter qu'il peut être difficile de respecter strictement la répartition des notes à attribuer en se référant à une pondération établie par rapport au programme d'études. Ceci est d'autant plus valable lorsque le nombre d'éléments d'information recueillis par l'épreuve est plutôt faible; et il en est ainsi pour plusieurs épreuves administrées en juin 2000 dans le cadre du CEP, nouveaux programmes. Pour être en mesure de respecter plus rigoureusement les paramètres fixés dans les tableaux de spécification, il faudrait donc s'assurer que les épreuves recueillent davantage d'éléments d'information; la couverture des champs de formation par les épreuves serait alors probablement plus complète. Cependant, une telle façon de procéder aurait un impact sur le format des épreuves et sur les coûts reliés à leur administration. Il importe alors de faire les choix qui respectent tant la rigueur de l'évaluation que les capacités d'administration du mécanisme.

Pertinence, fiabilité et validité des épreuves

Huit épreuves sont prévues pour l'examen du CEP, nouveaux programmes. Il y a eu quatre épreuves de français (expression écrite, lecture, orthographe, compréhension orale), une épreuve de mathématique, une épreuve de dessin, une épreuve d'éducation physique et sportive et une épreuve combinant les deux champs de l'éducation sociale et de l'éducation scientifique et technologique.

⁷ Il faut aussi noter qu'il y a eu rédaction de comptes-rendus par les secrétaires de l'atelier qui s'est déroulé à Pobè avant la venue du consultant et au cours de son séjour. Ces comptes-rendus seront disponibles chez le pilote du plan d'actions.

⁸ Il n'a pas été possible de traiter les données portant sur l'épreuve de dessin; il y avait alors une incohérence entre la base de données et les fichiers de saisie de l'information.

Nous avons pu noter quelques faiblesses par rapport aux épreuves utilisées en juin 2000 dans le cadre du CEP, nouveaux programmes. Les lacunes identifiées touchent parfois les épreuves elles-mêmes, souvent les grilles de correction. La présentation des *épreuves* remises aux élèves est soignée et il n'y a pas d'erreurs pouvant avoir une influence significative sur les réponses fournies par les élèves. Il y aurait toutefois lieu de mieux standardiser les *consignes* là où cela s'applique. Dans cet ordre d'idées, les Guides d'administration pourraient être intégrés, devenant alors un document contenant l'ensemble de l'information utile pour assurer une gestion harmonieuse de l'examen; un tel document n'aurait probablement pas à être repris d'une année à l'autre, à moins que des modifications aux procédures aient été apportées. Là où il y aurait des modifications à apporter concerne les *guides de correction*. L'ambiguïté de certaines consignes de notation a certainement eu un impact sur les résultats obtenus par les élèves. Il serait souhaitable d'uniformiser certains guides de correction, en particulier là où il y a des critères d'évaluation requérant l'attribution de notes suivant la technique de «crédits partiels».

Comme nous l'avons indiqué aux participants de l'atelier de Pobè, on peut classer les épreuves administrées en juin 2000 par rapport à cette classification : a) des épreuves écrites recueillant des éléments d'information fournis par des réponses des élèves à une série de questions; b) des tâches que l'on demande de réaliser et dont la production est évaluée à l'aide d'une grille de notation ou d'un schéma d'observation. Ordinairement, la première catégorie citée comprend des épreuves dont l'information peut être traitée plus facilement; on peut rattacher les épreuves suivantes à ce type d'épreuves : mathématique, éducation scientifique et technologique, éducation sociale, lecture. Les autres épreuves du CEP peuvent être considérées comme des tâches à réaliser par les élèves et dont la correction sera réalisée à l'aide d'une série de critères ou d'indicateurs de réussite. C'est ici qu'il faudrait concentrer ses efforts pour d'une part mieux préciser les critères à considérer et former les correcteurs à l'utilisation d'un tel outil d'évaluation.

Analyse statistique et pédagogique des résultats obtenus

Comme nous l'indiquions plus haut, des données ont été saisies sur un échantillon de copies d'élèves soumis aux épreuves du CEP, nouveau programme, de juin dernier. Environ vingt élèves par département ont été échantillonnés; ce sont les mêmes élèves qui ont été retenus pour l'ensemble des épreuves. L'objectif visé par cet exercice est d'examiner la qualité métrique de ces instruments. Il est vrai que la taille de l'échantillon aurait pu être plus grande; les données disponibles permettront quant même de disposer de tendances susceptibles de fournir des renseignements utiles pour se prononcer sur la qualité des épreuves utilisées et guider les concepteurs des futures épreuves. Nous avons retenu deux types de compilations statistiques : dans un premier temps, il y a eu analyses d'items afin d'examiner les qualités métriques des épreuves; par la suite, le calcul de résultats moyens par regroupement d'items correspondant aux tableaux de spécification a permis d'identifier les habilités et les contenus notionnels plus ou moins bien réussis.

Voici quelques constats faits à la suite de l'examen des tableaux statistiques. On a pu observer que le nombre d'éléments d'information présents dans les différentes épreuves varie considérablement (de un élément à 31). Il est entendu que plus il y a d'éléments d'information disponibles, plus fiables seront les indices statistiques rendus disponibles par des compilations statistiques. Certaines épreuves ont été très difficiles pour les élèves (Expression écrite, EST / ES, Mathématique); d'autres (EPS et Dessin) ont été très faciles. La cohérence interne des épreuves est acceptable là où l'on dispose d'un nombre suffisant d'éléments d'information (mathématique et éducation scientifique et technologique / éducation sociale). Quant à la discrimination des différents items ou critères, il y a généralement des lacunes qui mériteraient d'être examinées avec attention.

Les responsables de la DEC ont fournis au consultant des statistiques sur les sessions du CEP de 1994 à 2000. Nous pouvons dégager certaines tendances. Par rapport aux programmes intermédiaires, on constate une progression constante du nombre de candidats présents à l'examen du CEP (de près de 51000 en 1994 à près de 84000 en 2000). La proportion de candidats admis semble se stabiliser à 68%. Quant aux présences des candidats au CEP, nouveaux programmes, elles se situent à environ mille. Les taux d'admission affichent plus de 80%. On peut s'interroger sur les raisons expliquant de meilleurs succès des candidats provenant des nouveaux programmes si on les compare aux élèves des programmes intermédiaires. Est-ce que les épreuves seraient plus faciles? Est-ce que l'introduction des nouveaux programmes expliquent le succès des élèves? Il faudrait entreprendre des travaux de comparaison de l'instrumentation utilisée pour chaque format de CEP.

Coût et efficacité du nouveau format des épreuves

Juin 2000 marque la troisième administration du CEP, nouveaux programmes. Il y a eu, à chaque année, des activités préparatoires et des activités de suivi. C'est ainsi que des techniques pour réaliser les différentes opérations ont dû être arrêtées et maîtrisées par les nombreux partenaires impliqués : précision aux tableaux de spécification, élaboration des instruments (épreuves et grilles de correction), analyse de données recueillies auprès d'échantillons de candidats. La réglementation régissant cet examen a été révisée pour mieux tenir compte de la réalité et adapter le format à une généralisation qui arrive à grands pas; le souci des responsables demeure toutefois le respect de la rigueur et de l'équité du mécanisme d'évaluation à mettre en place pour remplir les responsabilités liées à la sanction des études primaires.

Même si les opérations prévues pour mettre sur pied l'examen 2000 du CEP, nouveaux programmes, ont été réalisées avec une maîtrise plus grande des techniques nécessaires, des améliorations aux pratiques d'élaboration de l'instrumentation ont pu être identifiées au cours des travaux réalisés à Pobè par les participants à cet atelier. La session de juin 2001 sera l'occasion d'évaluer concrètement la capacité des différents acteurs puisque le nombre de candidats sera beaucoup plus important si on compare à la situation des trois dernières sessions. Il y aura de grands défis à relever principalement par rapport à l'administration des épreuves et à leur correction.

Il semble sage de ne pas changer le format actuel du CEP, nouveaux programmes, tel que modifié après la tenue des sessions de 1998 et 1999 . La session de 2001 constituera un véritable test quant au format déjà arrêté. Il sera peut-être nécessaire d'apporter des changements à la suite de cette mini généralisation.

La session 2000 du CEP, nouveaux programmes

La préparation du matériel nécessaire à la tenue de la session 2000 du CEP a été réalisée avec un souci de mieux établir des liens avec les tableaux de spécification. C'est ainsi que des items ont été élaborés par rapport aux différentes dimensions à évaluer. Sans considérer la procédure retenue comme s'appuyant sur l'existence ou le développement de banques d'items, les travaux d'élaboration de l'instrumentation s'en sont largement inspirés. C'est ainsi qu'au terme des activités réalisées en mars 2000, un matériel d'évaluation respectant les paramètres des tableaux de spécification était disponible. Toutefois, il a été impossible de réaliser immédiatement l'assemblage de ce matériel en jeux d'épreuves. C'est une autre équipe qui a dû prendre le relais. Il serait souhaitable que l'ensemble des opérations soit finalisé par l'équipe mise sur pied pour l'élaboration de l'instrumentation; un groupe réduit pourrait alors prendre la relève pour réaliser un dernier examen du matériel proposé au directeur de la DEC.

Activités complémentaires réalisées au cours de la mission

Les termes de référence développés pour cette mission de consultation ont supposé la réalisation de plusieurs tâches qui ont dû se tenir en peu de temps. Toutefois, comme cela arrive souvent dans un tel contexte, des activités complémentaires furent ajoutées pour tenir compte de la situation du moment. Nous pouvons identifier trois catégories de travaux qui se sont réalisés en parallèle à ceux exigés par les termes de référence. Les voici rapidement présentés.

Au moment de la tenue des travaux de l'atelier de Pobè, il a été nécessaire d'ajouter au programme d'activités des *sessions de formation* sur les concepts reliés à l'évaluation des apprentissages. C'est ainsi que des mini sessions de formation à différentes techniques d'analyses statistiques ont été insérées au besoin. Le consultant, à la demande expresse des participants, a préparé rapidement un lexique présentant un ensemble de concepts sur l'évaluation des apprentissages.

Pour être en mesure de mettre à la disposition des participants des données statistiques qui ont apporté un éclairage complémentaire à l'analyse de l'instrumentation utilisée au CEP, nouveau programme, de juin dernier, il a été nécessaire d'exécuter des compilations statistiques à l'aide du logiciel utilisé depuis plus de cinq ans. La dernière version de ce programme informatique installé sur l'ordinateur utilisé à Pobè datait de plus de deux ans. Le consultant avait prévu installer une version plus récente de ce logiciel. Il a donc fallu prévoir une *mise à niveau des connaissances techniques* des personnes concernées afin de prendre en compte les modifications apportées au produit.

Au cours des deux dernières années, certains responsables du projet CLEF ont sollicité du consultant des renseignements sur la *mise en place d'un programme national d'évaluation des élèves*. Une telle activité avait eu lieu entre 1994 et 1996; elle a conduit à la réalisation d'évaluation d'échantillons de classes dans deux disciplines et la production de rapports qui ont été distribués auprès des responsables scolaires. Le sujet est encore revenu au cours de cette mission. Des rencontres ont été tenues pour préciser les activités réalisées et ce qui devrait être fait pour assurer une telle évaluation.

Suite des opérations

Les travaux réalisés par les participants à l'atelier de Pobè avant l'arrivée du consultant et au cours de son séjour ont permis d'une part de dresser un bilan de la session de juin 2000 du CEP et d'autre part de préparer les activités à prévoir pour la préparation du prochain CEP, nouveaux programmes. Nous

présentons ci-dessous un aperçu des opérations à réaliser et des suites qu'il est possible d'entrevoir.

- *Expertise nationale en techniques d'analyses du rendement des élèves.* Comme nous l'indiquons plus haut, il s'agit de la troisième édition de cette forme d'analyse du matériel et des procédures retenus pour la tenue du CEP, nouveaux programmes. La réalisation de ces examens successifs s'appuie sur des données recueillies auprès d'échantillons de candidats. Les participants aux différents ateliers organisés depuis 1998 ont pu acquérir des connaissances techniques reliées à l'analyse statistique adaptée à l'évaluation des apprentissages. Un équipe technique a développé l'expertise nécessaire à la réalisation des travaux exigés pour disposer de l'information requise. Ils sont de plus en plus nombreux les participants à maîtriser les concepts et les techniques permettant une analyse rigoureuse des résultats obtenus à des épreuves administrées dans le cadre du CEP.
- *Améliorations à apporter.* Même si on en est au troisième cycle d'analyse du CEP, nouveaux programmes, on constate qu'il y a encore des lacunes à corriger au matériel et aux procédures pour rendre cet examen plus rigoureux. Les analyses réalisées au cours du présent atelier a permis d'identifier des aspects à corriger pour disposer d'une instrumentation améliorée. Chaque équipe de travail a rédigé des rapports rendant compte des observations faites par les participants. Il faudra s'y référer au moment de l'élaboration des jeux d'épreuves nécessaires à la tenue du CEP de juin 2001.
- *Uniformisation de l'instrumentation.* Les ateliers qui se sont réalisés en 1998, 1999 et 2000 ont mis l'accent sur les travaux en équipes formées de spécialistes affectés à l'analyse d'un seul champ de formation. Ce mode de fonctionnement a permis un examen en profondeur de chaque champ de formation. On a toutefois constaté que les travaux ont évolué à des rythmes différents compte tenu de la composition des équipes et des particularités des champs de formation. Il est apparu important de réaliser des travaux regroupant des participants provenant des différents champs de formation. Il faudra tenter d'uniformiser certains aspects de l'instrumentation tout en étant conscient qu'il faut respecter les particularités des champs de formation.
- *Préparation de l'instrumentation 2001.* Il faudrait songer rapidement à la mise en place des ressources nécessaires à la préparation du matériel nécessaire à la tenue du CEP 2001. Il serait souhaitable de réserver assez de temps pour permettre à une équipe réduite de réaliser une révision finale de l'instrumentation en s'efforçant tout particulièrement d'uniformiser la présentation des différents documents. On pourrait tenter d'intégrer les différents guides d'administration dans le but d'éviter la redondance dans la diffusion de l'information.
- *Prochaines compilations touchant le CEP 2001.* À la suite de l'administration du CEP 2001, il semble nécessaire de prévoir une opération de recorection d'un échantillon de candidats, d'une entrée de ces données et de la réalisation de compilations statistiques. Nous l'avons déjà noté; il y aura beaucoup plus de candidats au CEP 2001, nouveaux programmes, en provenance d'un nombre plus important d'écoles. Nous proposons donc de retenir un échantillon plus important pour réaliser cette analyse; le nombre de candidats à retenir devrait être fixé à 300. Ceci permettra de réaliser des compilations plus fiables.
- *Disponibilité des ressources informatiques à la DEC.* Au cours de cet atelier, le personnel de la DEC a été beaucoup plus présent aux travaux d'analyse des instruments utilisés au CEP. D'un autre côté, une demande d'acquisition de matériel informatique a été formulée par cette Direction. Si on veut être en mesure de généraliser à moyen terme le format du CEP, nouveaux programmes, il sera important que la DEC dispose des outils nécessaires, à savoir l'équipement, les logiciels et les compétences techniques. Il faudrait examiner cette question dans un avenir très proche.

- *Ressources informatiques mises à la disposition du plan d'actions «Évaluation».* Le plan d'actions portant sur l'évaluation se voit confier de nombreux travaux techniques qui exigent le recours à des outils informatiques permettant l'exploitation des technologies supposant la disponibilité d'appareils performants. Les ordinateurs mis à la disposition du personnel de ce plan d'actions permettent de réaliser les travaux, il est vrai; mais, c'est souvent au prix de nombreuses difficultés. Il faudrait songer à renouveler certains appareils et y adjoindre des périphériques adaptés à ce champ d'expertise. Par ailleurs, des échanges d'information via le courrier électronique (par exemple, expédition de fichiers de données, de bases de données entre le pilote et le consultant) permettrait de diminuer les coûts puisque l'on pourrait réduire le nombre de séjours pour de la consultation sur place.

Léo Laroche
2 février 2001