

Etude scientifique de la qualité de la méthode de sélection des magistrats : analyse de la fidélité et de la validité de l'examen d'aptitude professionnelle du Conseil supérieur de la Justice

Emilie Michaux, Jean-Louis Michaux & Geert Vervaeke

1. Introduction : la sélection d'un personnel compétent

Un personnel qualifié représente un atout indispensable pour toute entreprise du secteur privé¹; les performances de ce personnel contribuent largement à assurer la compétitivité de l'entreprise². A contrario, un comportement préjudiciable adopté par certains membres du personnel, communément appelé « comportement contre-productif », risque de compromettre dangereusement le bon fonctionnement de cette organisation³.

Cette prise de conscience a conduit de nombreuses entreprises à investir dans une politique de ressources humaines de haute qualité apte à attirer et sélectionner un personnel compétitif⁴. Le rôle du département des ressources humaines est de dessiner le profil idéal d'un employé porteur des qualités performantes pour une activité spécifique et d'identifier le candidat qui correspond le mieux

¹ P.M. PODSAKOFF, S.B. MACKENZIE, N.P. PODSAKOFF et D.G. BACHRACH, "Scholarly influence in the field of management: a bibliometric analysis of the determinants of university and author impact in the management literature in the past quarter century", *Journal of Management* 2008, 641-720; R. SIMS, *Organizational Success through effective human resources management*, Westport, Quorum Books, 2002, 160; R.E. PLOYHART et T.P. MOLITERNO, "Emergence of the human capital resource: a multilevel model", *Academic Management Review* 2011, 127-150; P.M. WRIGHT et G.C. MCMAHAN, "Exploring human capital: putting human back into strategic human resource management", *Human Resource Management* 2011, 93-104

² A. VOMBERG, C. HOMBURG et T. BORNEMANN, "Talented people and strong brands: The contribution of human capital and brand equity to firm value", *Strategic Management Journal* 2014, 2122-2131; R. WILDEN, S. GUDERGAN et I. LINGS, "Employer branding: strategic implications for staff recruitment", *Journal of Marketing Management* 2010, 56-73

³ E. GRIJALVA et D. A. NEWMAN, "Narcissism and Counterproductive Work Behavior (CWB): Meta-Analysis and Consideration of Collectivist Culture, Big Five Personality, and Narcissism's Facet Structure", *Applied Psychology* 2015, 93-126; L. M. PENNEY et P.E. SPECTOR, "Narcissism and Counterproductive Work Behavior: Do Bigger Egos Mean Bigger Problems?", *International Journal of Selection and Assessment* 2002, 126-134; N.A. BOWLING et M.L. GRUYS, "Overlooked issues in the conceptualization and measurement of counterproductive work behavior", *Human Resource Management Review* 2010, 54-61

⁴ S.P. ROBBINS et M. COULTER, *Management. Eighth edition*, 2005, New Jersey, Pearson Education, p. 282

à ce profil⁵. A cette fin, les sélectionneurs font appel à des procédures soigneusement élaborées qui tentent d'évaluer les qualités nécessaires pour remplir une fonction déterminée⁶.

La recherche d'employés performants n'est pas l'apanage du secteur privé⁷. Les services publics, de même que le Service Public Fédéral de la Justice, ambitionnent cette même qualité de prestation de leurs services. Il en ressort que la qualité de la sélection de ce personnel joue un rôle clé pour le bon fonctionnement de l'organisation judiciaire. Outre la qualité du service rendu au citoyen, la Justice doit susciter la confiance de tout justiciable ce qui l'incite à reconnaître le bien-fondé de l'institution⁸. Une sélection de qualité contribue sans aucun doute à cette perception du citoyen⁹. Il va sans dire que la légitimité du système judiciaire, qui repose sur la reconnaissance du public, est une valeur primordiale à perfectionner en permanence¹⁰.

2. La sélection des magistrats en Belgique

A. Une évolution vers une sélection plus objective

La manière de sélectionner les magistrats en Belgique a subi déjà de nombreuses modifications¹¹. L'idée directrice de ces changements repose sur la recherche d'une sélection plus objective de ceux-ci¹². La création du collège de recrutement des magistrats en 1991 constitue un premier pas dans cette

⁵ A.M. RYAN et N.T. TIPPINS, "Attracting and selecting: What psychological research tells us", *Human Resource Management* 2004, 305-318; I.T. ROBERTSON et M. SMITH, "Personnel selection", *Journal of Occupational and organizational psychology* 2001, 441-472; J. F. SALGADO, C. VISWESVARAN et D. ONES, "Predictors used for personnel selection: an overview of constructs, methods and techniques", dans N. ANDERSON, D. ONES, H. K. SINANGIL et C. VISWESVARAN (eds.), *Handbook of Industrial, Work and Organizational Psychology. Volume 1: Personnel Psychology*, London, Sage Publications, 2001, p. 165-199; A. EVERS, N. ANDERSON et O.F. VOSKUIJL (eds.), *The Blackwell Handbook of Personnel Selection*, Oxford, England, Blackwell, 2005; E. DEROUX et F. DE FRUYT, "Developments in Recruitment and Selection Research", *International Journal of Selection and Assessment* 2016, 1-3

⁶ M. COOK, *Personnel selection: Adding value through people* (Fifth edition), Oxford, UK, Wiley-Blackwell, 2005; J.E. HUNTER et F.L. SCHMIDT, "Fitting people to jobs: The impact of personnel selection on national productivity", in M.D. DUNNETTE et E.A. FLEISMAN (Eds.), *Human performance and productivity: Human capability assessment*, Hillsdale, NJ, Lawrence Erlbaum Associates, 1982, 233-284

⁷ H. DE VRIES, V. BEKKERS et L. TUMMERS, "Innovations in the Public Sector: A Systematic Review and Future Research Agenda", Ottawa: IRSPM conference, 2014; J. GUTHRIE et J. DUMAY, "New frontiers in the use of intellectual capital in the public sector", *Journal of Intellectual Capital* 2015, 258-266

⁸ M. HOUGH, "How to measure trust? Survey measures of trust in justice" dans M. HOUGH et M. SATO (eds.) *Trust in justice: why it is important for criminal policy, and how it can be measured. Final report of the Euro-Justis project*, Helsinki, Academic Bookstore 2011, 18-24

⁹ S. MALENGREAU, "Montée en puissance et légitimité du juge", http://dev.ulb.ac.be/droitpublic/fileadmin/telecharger/theme_2/contributions/MALENGREAU-2-20070503.pdf; E. BROUILLET, « La légitimité fédérative du processus de nomination des juges à la Cour suprême du Canada », *RGD* 2011, 279-293 ; S. TURENNE, *Fair reflection of society in judicial systems. A comparative study*, New York, Springer, 2015, 236p.

¹⁰ E. MICHAUX, G. VERVAEKE, D. VAN DAELE, et R. MORTIER, "De rechter ten tijde van het New Public Management: een verhaal van onverzoenbare verwachtingen", *RW* 2018, 923-940

¹¹ M. VAES, G. VERVAEKE, R. MORTIER, D. VAN DAELE et D. BARELDS, "Naar een evaluatie van de selectie van magistraten", *RW* 2014-2015, 563-573

¹² R. ANDERSEN, « La nomination des juges en Belgique », *RGD* 2006, 689-708

évolution¹³. Le rôle de ce collège est de mettre un terme à la sélection dite « politique » qui prenait un rôle prépondérant dans le choix des magistrats par rapport aux valeurs intrinsèques des candidats¹⁴. Le collège de recrutement des magistrats a introduit un double accès à la magistrature : pour les juristes moins expérimentés, par la création d'un concours d'admission à un stage judiciaire, pour des juristes déjà expérimentés, par la possibilité de participer à un examen de sélection qui autorise de postuler une place de magistrat¹⁵.

L'affaire Dutroux et ses inquiétantes retombées dans le domaine judiciaire ouvrit une nouvelle étape dans l'objectivation de la sélection des magistrats. Pour restaurer la confiance des citoyens dans la justice et limiter davantage l'influence de l'appartenance à un parti politique dans le recrutement et la nomination des magistrats¹⁶, il fut décidé d'instaurer un Conseil supérieur de la Justice (CSJ) dont la création remonte à 1999. Le CSJ, qui remplace le collège de recrutement des magistrats, est inscrit dans la constitution belge¹⁷. Cet organisme est responsable de la sélection des futurs magistrats par le biais de sa commission de nomination et de désignation (CND). Cette commission, composée de 14 membres du CSJ avec une parité entre magistrats et non-magistrats, est chargée d'organiser les examens d'accès à la magistrature et de présenter des candidats à une nomination ou à une désignation. Dans un premier temps, le CSJ a maintenu le double accès à la magistrature, mais, en 2005, a introduit l'examen oral d'évaluation réservée aux candidats qui disposent d'au moins 20 années d'expérience professionnelle dans le domaine juridique que l'on dénomme la « troisième voie »¹⁸.

B. L'examen d'aptitude professionnelle: une voie d'accès à la magistrature pour des juristes expérimentés

Dans cette étude, notre contribution se limite à l'examen d'aptitude professionnelle et n'aborde pas le concours d'admission au stage judiciaire ni l'examen oral d'évaluation. Nous présentons donc uniquement une description détaillée de l'examen d'aptitude professionnelle.

¹³ Loi du 18 juillet 1991 modifiant les règles du Code judiciaire relatives à la formation et au recrutement des magistrats, *MB* 26 juillet 1991

¹⁴ J.E. KRINGS, « *La formation et le recrutement des magistrats* », Bruxelles : Ministère de la Justice, 1976, 124p.

¹⁵ Arrêté Royal du 5 mars 1992 relatif au recrutement des magistrats; Arrêté ministériel du 29 octobre 1992 portant ratification des programmes d'examen concernant le concours d'admission au stage judiciaire et de l'examen d'aptitude professionnelle, *MB* 31 octobre 1992; Loi du 6 août 1993 portant des dispositions sociales et diverses, *MB* 9 août 1993, p. 17.876

¹⁶ C. SÄGESSER, « Législatif, exécutif et judiciaire. Les relations entre les trois pouvoirs », *Dossiers du CRISP* 2016, 9-71

¹⁷ Loi du 20 novembre 1998 modifiant la Constitution, *M.B.* 24 novembre 1998, 37.679 ; Loi modifiant certaines dispositions de la deuxième partie du Code judiciaire concernant le Conseil supérieur de la Justice, la nomination et la désignation des magistrats et instaurant un système d'évaluation pour les magistrats, *M.B.*, 2 février 1999.

¹⁸ A.M. du 4 mai 2006 portant ratification du programme de l'examen oral d'évaluation, *M.B.* 12 mai 2006; . DE GROOTE, G. CLEEMPUT et G. VERVAEKE, "Het mondelinge evaluatie-examen: waar blijven de witte raven?", *Ad Rem* 2011, 21-33

Conformément aux dispositions légales, le ministre de la Justice charge la commission de nomination et de désignation du CSJ de préparer le programme d'examen¹⁹. Après approbation de ce programme par l'assemblée générale du CSJ, il est envoyé au ministre de la Justice pour ratification et publication au Moniteur belge²⁰. Par la suite, est lancé par le Moniteur un appel à candidatures²¹, recevables dans le mois suivant la publication²². Les conditions de recevabilité portent sur un curriculum vitae selon un formulaire type établi par le CSJ, une copie du diplôme de licencié ou de master en droit, le numéro de registre national de l'Etat belge et une attestation d'une expérience juridique d'au moins quatre ans au cours des cinq dernières années. Le candidat est amené à faire un choix de la matière du droit dans laquelle il est interrogé, à savoir le droit civil (y compris le droit judiciaire), le droit pénal (y compris la procédure pénale) ou le droit social (y compris le droit judiciaire). Tous les candidats qui remplissent les conditions susmentionnées sont convoqués par le CSJ à la sélection. Récemment il fut légiféré que les candidats ayant échoué cinq fois à l'examen étaient exclus de toute participation ultérieure²³.

i. Techniques et procédures de sélection de l'examen d'aptitude professionnelle

Chaque méthode de sélection repose sur des techniques et sur un processus de sélection. Les techniques font référence aux tests de sélection. Le processus concerne la manière dont ces techniques sont appliquées, telles que la séquence et la combinaison de tests, la manière ou l'ordre de leur application. Depuis la création du CSJ, plusieurs techniques et processus se sont succédé dans la méthode de sélection des magistrats. Ainsi, on distingue dix techniques: la rédaction d'un jugement, l'analyse d'un arrêt de la Cour de cassation, la dissertation, le questionnaire à choix multiple, la résolution d'un cas pratique, le test cognitif-analytique, le questionnaire de personnalité, l'exposé critique, la discussion à partir d'un casus et la discussion sur la résolution d'un cas pratique.

La « rédaction d'un jugement » est une épreuve écrite qui s'effectue sur base d'un dossier complet mis à la disposition des candidats, et porte sur une cause de droit civil, droit pénal ou droit social selon le choix du candidat lors de l'inscription. Les candidats disposaient originellement de quatre heures pour rédiger ce jugement correct tant au niveau de la forme que du fond²⁴. Ultérieurement le temps agréé fut de cinq²⁵ et même six heures²⁶.

¹⁹ Art. 259bis-9, §3 C. Jud.; Art. 244, 2°, premier paragraphe de la loi du 6 juillet 2017 Loi portant simplification, harmonisation, informatisation et modernisation de dispositions de droit civil et de procédure civile ainsi que du notariat, et portant diverses mesures en matière de justice (1), *M.B.* 24 juillet 2017

²⁰ Art. 259bis-9, §3 C. Jud.

²¹ Art. 2 A.R. 21 septembre 2000 déterminant les modalités et les conditions d'organisation de l'examen d'aptitude professionnelle et du concours d'admissions au stage judiciaire, *M.B.* 30 septembre 2000, 33.454

²² Art. 259octies, §1, C. Jud.; Art. 4 A.R. 21 septembre 2000 déterminant les modalités et les conditions d'organisation de l'examen d'aptitude professionnelle et du concours d'admissions au stage judiciaire, *M.B.* 30 septembre 2000, 33.454

²³ Art. 244, 2°, premier paragraphe de la loi du 6 juillet 2017 Loi portant simplification, harmonisation, informatisation et modernisation de dispositions de droit civil et de procédure civile ainsi que du notariat, et portant diverses mesures en matière de justice (1), *M.B.* 24 juillet 2017

²⁴ A.M. du 7 août 2001 portant ratification du programme de l'examen d'aptitude professionnelle et le concours d'admission au stage judiciaire, *M.B.* 9 août 2001

²⁵ A.M. du 19 juin 2003 portant ratification du programme de l'examen d'aptitude professionnelle et le concours d'admission au stage judiciaire, *M.B.* 1 septembre 2003

²⁶ A.M. du 23 septembre 2010 portant ratification du programme de l'examen d'aptitude professionnelle et le concours d'admission au stage judiciaire, *M.B.* 29 septembre 2010

L'«analyse d'un arrêt de la Cour de cassation» impose aux candidats d'examiner un jugement de la Cour de cassation dans un langage accessible portant sur : les faits pertinents de l'affaire, les moyens de cassation, la disposition légale violée, les décisions et motifs critiqués de la décision attaquée et griefs, la décision de la Cour de cassation et ses attendus, et les conséquences et la portée de l'arrêt. La durée de cette épreuve écrite est de quatre heures²⁷.

Pour «la dissertation» sur un sujet d'actualités imposé à caractère social, économique, politique ou culturel en relation avec le droit, le temps imparti a varié de deux²⁸, trois²⁹ à quatre³⁰ et cinq³¹ heures et la rédaction est limitée à quatre pages.

Le «questionnaire à choix multiple» vise à évaluer les connaissances juridiques de base, les connaissances sur le fonctionnement de l'organisation judiciaire et les connaissances générales des candidats. Le questionnaire à choix multiples est toujours proposé en combinaison avec la dissertation. Les candidats disposent de cinq heures trente au total pour ces deux épreuves³².

Pour l'épreuve écrite «la résolution d'un cas pratique», les candidats sont amenés à formuler une solution juridique sur la base d'un dossier de pièces, en tenant compte de manière optimale de la spécificité des faits ainsi que du contexte particulier et sociétal qui caractérise le casus³³. L'intention est de tester le pragmatisme juridique des candidats³⁴. Les candidats sont invités à évaluer la réalité dynamique du cas et à justifier l'approche juridique retenue à partir de cette réalité³⁵. Les candidats optent pour une matière de droit (civil, pénal ou social) et disposent de cinq heures pour résoudre l'exercice.

Le «test cognitif-analytique» vise à mesurer la capacité à résoudre des problèmes, à formuler et à appliquer des connaissances et à s'adapter à l'environnement grâce à l'apprentissage. Ce test permet d'évaluer le pouvoir de raisonnement abstrait et l'intelligence verbale.

Le «questionnaire de personnalité» mesure les neuf compétences génériques de base qui, selon le CSJ, sont déterminantes pour l'exercice de la fonction de magistrat à savoir : l'esprit de décision, l'intégrité, la faculté d'adaptation, la capacité à planifier, à organiser et à établir des priorités, l'empathie et la sociabilité, la résistance au stress, la collaboration et la collégialité, la dominance / la gestion appropriée du pouvoir, l'auto-réflexion.

²⁷ A.M. du 10 septembre 2002 portant ratification du programme de l'examen d'aptitude professionnelle et le concours d'admission au stage judiciaire, *M.B.* 17 septembre 2002

²⁸ A.M. du 7 août 2001 portant ratification du programme de l'examen d'aptitude professionnelle et le concours d'admission au stage judiciaire, *M.B.* 9 août 2001

²⁹ A.M. du 28 septembre 2000 portant ratification du programme de l'examen d'aptitude professionnelle et le concours d'admission au stage judiciaire, *M.B.* 30 septembre 2000

³⁰ A.M. du 24 août 2004 portant ratification du programme de l'examen d'aptitude professionnelle et le concours d'admission au stage judiciaire, *M.B.* 15 septembre 2004

³¹ A.M. du 19 juin 2003 portant ratification du programme de l'examen d'aptitude professionnelle et le concours d'admission au stage judiciaire, *M.B.* 1 septembre 2003

³² A.M. du 10 juillet 2012 portant ratification du programme de l'examen d'aptitude professionnelle et le concours d'admission au stage judiciaire, *M.B.* 3 septembre 2012

³³ A.M. du 20 août 2015 portant ratification du programme de l'examen d'aptitude professionnelle et le concours d'admission au stage judiciaire, *M.B.* 7 septembre 2015

³⁴ R. POSNER, *How Judges Think*, Harvard University Press, Cambridge, 2008

³⁵ F. FLEERACKERS, "Rechtsdenken, pragmatisme en juridische intuïtie", dans F. FLEERACKERS et R. VAN RANSBEECK (eds.) *Juristen over rechtsdenken*, Brussel, Larcier, 2015, p.15

Lors de « l'exposé critique », les candidats disposent de dix minutes pour présenter oralement un sujet de leur choix portant sur le droit judiciaire privé ou la procédure pénale, suivi d'un échange de vues avec le jury sur les tests écrits³⁶.

Pour la « discussion à partir d'un casus », les candidats disposent de 60 minutes de préparation pour étudier un cas juridique. Par la suite, ils présentent oralement leur analyse à un jury et répondent à ses questions. Le cas échéant, un échange de vues est organisé sur d'autres questions juridiques générales, le fonctionnement de l'organisation judiciaire, les autres épreuves écrites, les résultats des tests psychologiques et la motivation du candidat³⁷.

La « discussion sur la résolution du cas pratique » est toujours présentée conjointement avec l'épreuve écrite « la résolution d'un cas pratique ». Lors de cette épreuve, les candidats élaborent oralement les résultats de leur épreuve écrite et répondent aux questions du jury à ce sujet. Suit un échange de vues sur d'autres questions juridiques générales, le fonctionnement de l'organisation judiciaire, le statut et la déontologie du magistrat, les résultats des tests psychologiques, la motivation et les compétences du candidat³⁸.

ii. Les changements dans l'examen d'aptitude professionnelle

Au fil des années, le CSJ a révisé à de nombreuses reprises le processus de sélection des magistrats. Ces changements démontrent une volonté de valoriser cette sélection³⁹. C'est en effet un devoir moral pour le CSJ de remettre en question les méthodes de sélection choisies, d'optimiser ce choix et de rendre compte des méthodes de sélection choisies⁴⁰.

A l'analyse de l'évolution du processus de l'examen d'aptitude professionnelle, nous distinguons trois périodes (cf. fig. 1). La première période (2000-2002) se caractérise par une double épreuve écrite, suivie d'une épreuve orale. Tous les candidats présentent les deux épreuves écrites : la dissertation et la rédaction d'un jugement. Seuls les lauréats, c'est-à-dire les candidats qui obtiennent un score d'au moins 12/20 pour l'ensemble des épreuves écrites, sont admis à l'épreuve orale qui consiste en « l'exposé critique ». Les lauréats de l'exposé critique sont autorisés à se porter candidats à un poste de magistrat.

La deuxième période (2002-2012) conserve la double épreuve écrite mais introduit le système de 'cascade'. Durant cette deuxième période, la première épreuve écrite est éliminatoire pour les candidats dont la cote est inférieure à 12/20 ; il en est de même pour la seconde épreuve écrite qui

³⁶ A.M. du 28 septembre 2000 portant ratification du programme de l'examen d'aptitude professionnelle et le concours d'admission au stage judiciaire, *M.B.* 30 septembre 2000

³⁷ A.M. du 10 septembre 2002 portant ratification du programme de l'examen d'aptitude professionnelle et le concours d'admission au stage judiciaire, *M.B.* 17 septembre 2002

³⁸ A.M. du 11 juillet 2017 portant ratification du programme de l'examen d'aptitude professionnelle et le concours d'admission au stage judiciaire, *M.B.* 5 septembre 2017

³⁹ . MORTIER, "Rechtsdenken zoals het zich uit bij de examens en de voordrachten voor benoeming van de Hoge Raad voor de Justitie", dans F. FLEERACKERS et R. VAN RANSBEECK (eds.) *Juristen over rechtsdenken*, Brussel, Larcier, 2015, p.18

⁴⁰ C. DE GROOTE, G. CLEEMPUT et G. VERVAEKE, "Het mondelinge evaluatie-examen: waar blijven de witte raven?", *Ad Rem* 2011, 21-33

reste éliminatoire dans les mêmes conditions. Une épreuve orale suit cette seconde épreuve écrite. Lors de cette deuxième période, l'épreuve orale consiste dans « la discussion à partir d'un cas. » Seuls les candidats qui obtiennent un score supérieur à 12/20 à l'ensemble de l'examen sont autorisés à postuler une place vacante de magistrat.

Le contenu des épreuves écrites de cette deuxième période a subi de nombreuses modifications dont nous abordons une analyse plus détaillée. Comme le montre la figure 2, la deuxième période se divise en cinq sous-périodes selon le contenu et la séquence des épreuves écrites. La rédaction d'un jugement suivi par l'analyse d'un arrêt n'est organisée qu'une seule année (en 2002), de même que la séquence inverse l'année suivante (en 2003). Ensuite pendant deux années (en 2004 et en 2005) la première épreuve est la rédaction d'un jugement et la seconde, la dissertation. La séquence est inversée pendant les trois années suivantes (en 2006, 2007 et 2008). Enfin à partir de 2009 jusqu'à 2012, un questionnaire à choix multiple (QCM) est ajouté à la première épreuve écrite. La première épreuve écrite comprend alors deux sous-épreuves : le QCM et la dissertation. Ce sont les candidats qui obtiennent un score d'au moins 12/20 pour l'ensemble des deux épreuves, qui ont le droit de présenter la seconde épreuve écrite : la rédaction d'un jugement. Depuis 2010, les tests psychologiques, cognitif-analytique et de personnalité, font partie des épreuves imposées aux lauréats des épreuves écrites.

La troisième période (2012-2018) est marquée par la résolution d'un cas pratique, épreuve écrite unique, éliminatoire pour l'épreuve orale. A nouveau le seuil de 12/20 est appliqué pour l'admission à l'épreuve orale qui est constituée de la « discussion sur la résolution du cas pratique ».

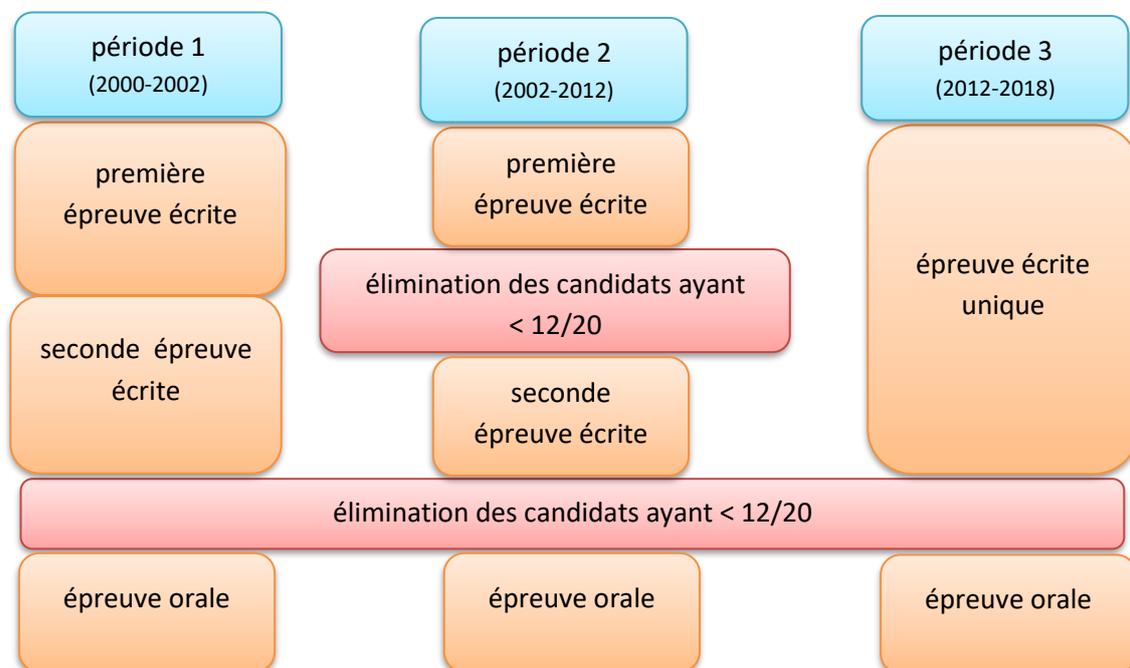


Fig.1. Schéma de l'évolution du processus de l'examen d'aptitude selon les trois périodes

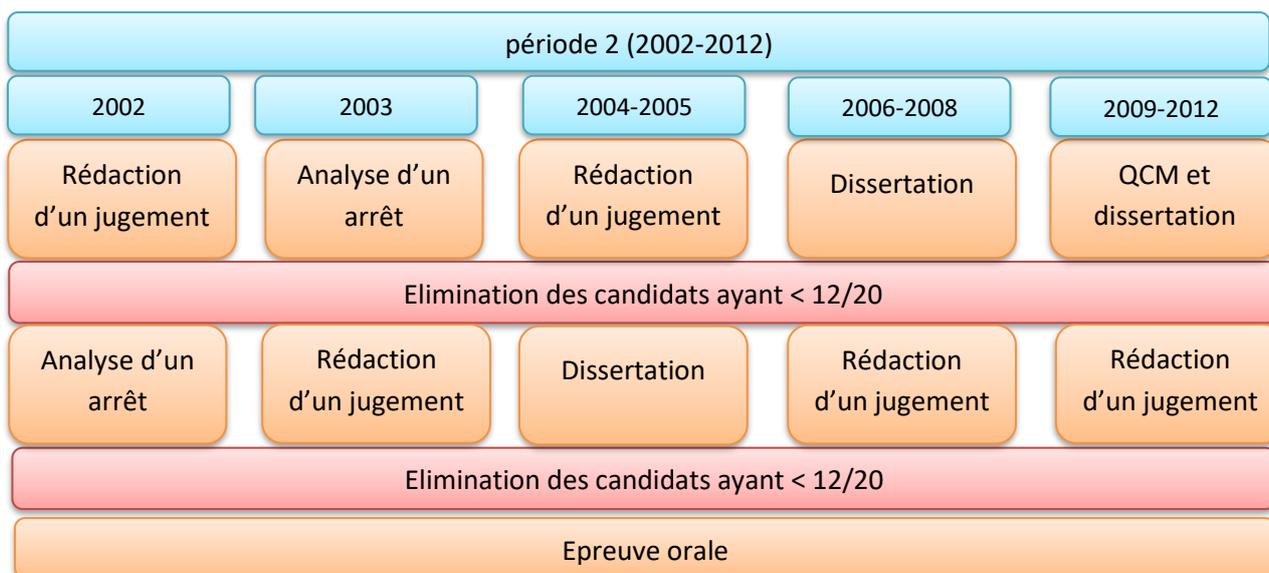


Fig. 2. Schéma des sous-périodes de la deuxième période regroupé par contenu des épreuves écrites

iii. L'évaluation de l'examen d'aptitude professionnelle

L'évaluation des épreuves écrites (mis à part le QCM) et de l'épreuve orale⁴¹ est confiée à un jury composé de huit membres de la commission de nomination et de désignation⁴². La composition du jury change chaque année mais veille à respecter une parité professionnelle (magistrats – non-magistrats). Cependant, le choix des membres du jury n'est nullement systématisé ni justifié.

L'évaluation des épreuves écrites (sauf le QCM) est entièrement «aveugle». Cela signifie non seulement que les membres du jury ne connaissent pas l'identité des candidats qu'ils évaluent, mais également que les candidats ne peuvent en aucune manière révéler leur identité dans leur réponse aux épreuves écrites sous peine d'exclusion. Depuis 2013, les épreuves écrites sont réalisées sur ordinateur pour rendre impossible la reconnaissance de l'écriture et pour garantir une meilleure objectivité.

Chaque membre du jury évalue chaque candidat pour chacune des épreuves, écrite et orale. Chaque candidat obtient, donc, par épreuve une cote moyenne de l'ensemble des examinateurs. Ce score moyen peut encore être majoré pour l'ensemble des candidats après une délibération du jury d'examen⁴³.

La correction du QCM se fait de façon électronique. L'examen comporte trente questions avec quatre alternatives de réponse. Une réponse correcte donne droit à 1 point. Une question qui n'a pas reçu de

⁴¹ A cette fin, les membres du jury sont formés dans l'entretien de sélection semi-directif.

⁴² Le nombre exact de membres du jury varie entre sept en 2008 et dix en 2010. Depuis 2012 deux sous-groupes sont formés selon la spécialité de droit avec entre quatre (en 2016) et six membres du jury.

⁴³ P. DETROZ, V. CRAHAY, G. JEANMART et A. PIAZZA, *Examens d'accès à la magistrature. Evaluations des procédures de sélection. Rapport d'audit, 2012, ULG, 70*

réponse est cotée 0. Une réponse incorrecte est pénalisée par la soustraction d' 1/3 de point. Le score de l'épreuve est rapporté à un total de 20.

Le test cognitif-analytique et le questionnaire de personnalité sont réalisés et évalués par des psychologues désignés par la commission de nomination et désignation du CSJ⁴⁴. Les résultats du questionnaire de personnalité est validé lors d'un entretien avec le candidat⁴⁵. Les psychologues mandatés fournissent au jury d'examen un rapport écrit sur les résultats des tests psychologiques ainsi que sur les résultats de l'entretien, mais ne font pas partie du jury d'examen. Durant les cinq premières années de leur installation, les résultats du test cognitif-analytique et du questionnaire de personnalité ne conduisaient, en aucun cas, à l'exclusion d'un candidat. Depuis 2015, ces résultats sont à même de motiver l'élimination d'un candidat⁴⁶.

iv. Le suivi de l'examen d'aptitude professionnelle

Les lauréats de l'examen, c'est-à-dire ceux qui ont obtenu une cote de 12/20 ou plus à la dernière épreuve, ont le droit de se porter candidats à un poste vacant de magistrat. Ces vacances de poste sont publiées par le ministre de la Justice dans le Moniteur belge⁴⁷. Les candidats sont auditionnés par la commission de nomination et de désignation lors d'un entretien d'embauche, qui donne ensuite son aval pour une proposition de nomination sur la base d'un accord de deux tiers des membres de cette commission. Il appartient au ministre de la Justice de nommer ou non le candidat en motivant son refus⁴⁸.

3. La qualité de la méthode de sélection

L'évolution vers une sélection plus objective des magistrats est sans aucun doute un développement opportun à promouvoir et à encourager. Cependant on ignore dans quelle mesure la méthodologie contribue à la sélection des meilleurs magistrats et si, réellement, celle-ci aide à évaluer la « maturité et la capacité nécessaires à l'exercice de la fonction de magistrat ». Après tout, les méthodes de sélection actuelles ne fournissent pas de réponse définitive à la question de savoir si les objectifs fixés dans le processus de sélection sont effectivement atteints. Cette validation peut uniquement se faire par une évaluation systématique, indépendante et scientifique. Une telle validation, bien que nécessaire, n'a jusqu'à présent, et à notre connaissance, jamais été effectuée pour la sélection des magistrats ni en Belgique, ni dans d'autres pays.

⁴⁴ *idem*

⁴⁵ A.M. du 11 juillet 2017 portant ratification du programme de l'examen d'aptitude professionnelle et le concours d'admission au stage judiciaire, *M.B.* 5 septembre 2017

⁴⁶ CSJ, *Rapport Annuel* 2015, http://www.hrij.be/sites/default/files/press_publications/20160511-jaarverslag2015-fr.pdf, p.7

⁴⁷ Art. 244, 2°, premier paragraphe de la loi du 6 juillet 2017 Loi portant simplification, harmonisation, informatisation et modernisation de dispositions de droit civil et de procédure civile ainsi que du notariat, et portant diverses mesures en matière de justice (1), *M.B.* 24 juillet 2017

⁴⁸ Art. 259ter. C. Jud.

A. Paramètres scientifiques de qualité: la fidélité et la validité

Une évaluation scientifique de la méthode de sélection vérifie la qualité de celle-ci sur la base de deux "paramètres de qualité": la fidélité et la validité. Seules les méthodes de sélection à la fois fidèles et valides atteignent l'objectif recherché. La fidélité d'une méthode de sélection, dans le domaine des sciences sociales, fait référence à la mesure par laquelle cette méthode conduit au même résultat lorsqu'elle est répétée à un autre moment par un autre chercheur⁴⁹. Cela montre à quel point les résultats obtenus sont exempts d'erreurs aléatoires et de biais non pertinents, et sont indépendants de la personne examinatrice, du lieu ou du moment de l'enquête⁵⁰. En d'autres termes, la fidélité est une mesure de la cohérence et de la stabilité d'une méthode de sélection. La validité est un indicateur de la mesure dans laquelle le résultat d'une méthode de sélection est conforme à la réalité⁵¹. Elle répond donc à la question de savoir si « la méthode mesure réellement ce qu'elle comptait mesurer » et si aucune erreur systématique et régulière ne fausse les résultats.

La fidélité est une condition nécessaire mais non suffisante pour la validité. Une méthode de sélection, visant à estimer la future performance du candidat dans son travail, effectuée à plusieurs reprises, peut en effet obtenir les mêmes résultats, et donc être fidèle, sans pour cela refléter la performance future au travail du candidat. D'un autre côté, il n'est pas possible que le résultat d'une méthode de sélection reflète correctement la performance professionnelle future d'un candidat lorsque plusieurs évaluateurs obtiennent des résultats différents avec la même méthode. Il est donc important de déterminer d'abord le degré de fidélité d'une méthode avant de déclarer sa validité.

On distingue différentes formes de fidélité et de validité⁵². Concernant la fidélité, on distingue sa stabilité dans le temps (test-retest), selon les examinateurs (fidélité inter-juge), et sa cohérence interne (convergence de diverses questions visant à mesurer le même concept). La validité se décompose en validité interne et externe. La validité interne, qui cherche à savoir si l'on mesure ce qu'on souhaite mesurer, est divisée en validité apparente (« est-ce qu'une méthode semble mesurer ce qu'elle souhaite mesurer »), validité du contenu (« dans quelle mesure une méthode est représentative du ou des construits mesurés et de leurs différentes facettes »), validité conceptuelle (« est-ce que la méthode mesure vraiment le ou les construits pour lesquels elle a été conçue et offre-t-elle une mesure adéquate du modèle théorique sur lequel elle s'appuie ») et validité prédictive (« les résultats d'une méthode peuvent-ils réellement prédire des comportements futurs pertinents »). La validité externe est l'aptitude d'une méthode à permettre de généraliser les informations recueillies au-delà des manifestations observées. Ce qui veut dire que cette validité externe est un indicateur de la possibilité de généraliser des cas particuliers observés à l'ensemble des cas observables.

⁴⁹ E. MICHAUX, E. JASPAERT, et G. VERVAEKE, "Niets is wat het lijkt. Resultaten van empirisch onderzoek in het recht", dans F. FLEERACKERS et R. VAN RANSBEECK (eds.) *Recht en Empirie*, Brussel, Larcier, 2017, p.79

⁵⁰ A. GROUS, *Eléments d'analyse de la fiabilité et du contrôle de qualité : statistiques appliquées par l'exemple*, Paris, Lavoisier, p.213

⁵¹ L.R. GOMEZ-MEJIA, D.B. BALKIN et R.L. CARDY, *Managing human resources*, Englewood Cliffs, NJ, Prentice Hall, 2009, p. 125

⁵² R. LECOMTE et L. RUTMAN, *Introduction aux méthodes de recherche évaluative*, Québec, Les presses de l'Université Laval, 1982

B. Objectif de cette contribution : analyse de la qualité de l'examen d'aptitude professionnelle de la section néerlandophone

Cette contribution tente d'examiner la qualité de l'examen d'aptitude professionnelle du secteur néerlandophone depuis la création du Conseil supérieur de la Justice par une analyse de la fidélité inter-évaluateur et de la validité conceptuelle. Le choix se porte sur ces deux types de fidélité et validité pour différentes raisons.

Premièrement, la cote octroyée aux candidats est basée sur l'évaluation de l'aptitude professionnelle par différents examinateurs. Il est donc important d'étudier si les différents membres du jury ont tendance à évaluer un même candidat de la même façon. Finalement, le but d'inclure plusieurs examinateurs est de rendre l'évaluation plus objective et d'éliminer au maximum l'influence des caractéristiques personnelles de l'évaluateur. Le type de fidélité qui recherche à identifier cette forme d'objectivité, est la fidélité inter-évaluateur. Ce type de fidélité fait référence au degré d'accord entre les divers évaluateurs utilisant le même instrument pour évaluer la même situation⁵³. Par ailleurs, plusieurs chercheurs ont démontré que la fidélité inter-évaluateur pour l'appréciation de la performance professionnelle est généralement plutôt faible ce qui risque de fausser les résultats des méthodes de sélection⁵⁴. Il est donc nécessaire de rechercher l'existence de ce type de fidélité.

La fidélité inter-évaluateur est étudiée de deux façons⁵⁵ : tout d'abord, par l'analyse d'une différence entre les scores moyens attribués par les différents examinateurs. Il y a lieu de repérer si un ou plusieurs évaluateurs sont systématiquement plus indulgents ou plus exigeants dans leur jugement. Pour ce faire, on réalise l'analyse simple de la variance qui établit l'existence d'écarts statistiquement significatifs entre les cotes moyennes de tous les évaluateurs pour la même épreuve. Un écart statistiquement significatif est une déviation dont on sait avec au moins 95% de certitude qu'elle n'est pas due au hasard. Seules les différences statistiquement significatives sont rapportées. Ensuite, par l'évaluation de la concordance inter-évaluateur. Il importe d'analyser dans quelle mesure les différents correcteurs ont tendance à évaluer une même copie de manière comparable. De ce fait, on étudie si le classement des différents candidats est similaire pour les différents examinateurs. Afin de pouvoir analyser ce sujet, nous procédons au calcul des coefficients intra-classe⁵⁶. Le coefficient intra-classe (ICC) est un score entre 0 et 1 qui représente le pourcentage de variation dans les évaluations réellement due aux variations dans les prestations des candidats. La concordance est d'autant plus importante que le coefficient est proche de 1. Une concordance parfaite, rarement atteinte, n'est pas nécessairement recherchée. Il existe plusieurs modèles d'interprétation du coefficient. En fonction de l'objectif recherché : scientifique vs. pratique, une autre valeur de seuil existe pour parler d'une

⁵³ C. PISCICELLI, J. BARRA, P. DAVOINE, A. CHRISPIN, S. NADEAU et D. PERENNOU, « Fidélité inter- et intra-évaluateur de la verticale visuelle chez les patients AVC », *Neurophysiologie Clinique* 2015, 412

⁵⁴ M.B. HARARI et C. VISWESVARAN, "Individual job performance", dans D.S. ONES, N. ANDERSON, C. VISWESVARAN et H.K. SINANGIL (red.), *The SAGE handbook of industrial, work and organizational psychology. 2nd Edition*, London, SAGE Publications Inc., 2018, 55-72

⁵⁵ N. GISEV, J.S. BELL, T.F. CHEN, "Interrater agreement and interrater reliability: Key concepts, approaches, and applications", *Research in Social and Administrative Pharmacy* 2013, 330-338

⁵⁶ ⁵⁶ P.E. SHROUT et J.L. FLEISS, "Intraclass correlations: Uses in assessing rater reliability", *Psychological Bulletin* 1979, 420-428; K.L. GWET, *Handbook of inter-rater reliability, 4th Edition*, Gaithersburg, Advanced Analytics, LLC, 2014, p.192

« bonne concordance »⁵⁷. Pour des décisions importantes, comme la nomination et sélection, une valeur minimale de .80 est appliquée comme mesure d'une bonne concordance inter-évaluateur.⁵⁸

La validité conceptuelle et la validité prédictive sont considérées comme les formes de validité les plus importantes pour la sélection du personnel⁵⁹. Cependant, avant de pouvoir se prononcer sur la validité prédictive, il faut s'assurer que la méthode adoptée mesure vraiment le ou les construits pour lesquels elle a été conçue (la validité conceptuelle). Généralement la validité conceptuelle est déterminée par l'examen du degré de cohérence entre les résultats de la méthode utilisée et les résultats d'autres méthodes validées mesurant le même construit ou d'autres construits, liés mais non similaires. Devant une forte corrélation, c'est-à-dire un haut degré de cohérence entre les scores des tests censés mesurer le même construit et la faible corrélation entre les scores des tests qui ne sont pas censés mesurer la même chose, on parle d'une grande validité conceptuelle. Une autre façon d'identifier la validité conceptuelle consiste à voir si certains facteurs, qui théoriquement ne devraient pas modifier le résultat d'une méthode, influencent tout de même les performances de cette méthode. En d'autres termes, on recherche l'existence d'un biais systématique dans les résultats du test, biais lié à des facteurs qui ne devraient pas jouer de rôle. Cette contribution examine si certains facteurs non pertinents influencent systématiquement les chances de réussite à l'examen d'aptitude professionnelle. L'influence des facteurs non pertinents suivants sur la performance des candidats fait l'objet de notre contribution à savoir le sexe, l'âge, l'université, le grade académique et le choix de la matière du droit⁶⁰. En fonction du facteur étudié, différentes statistiques sont appliquées.⁶¹ A nouveau, uniquement les différences statistiquement significatives sont rapportées.

4. Résultats des analyses de la fidélité inter-évaluateur

A. Fidélité inter-évaluateur

Les analyses démontrent que pour chaque année et pour chaque épreuve écrite, c'est-à-dire pour la correction de la rédaction d'un jugement, de l'analyse de l'arrêt, de la dissertation et de la résolution du cas pratique, des différences significatives existent entre les correcteurs. Cela signifie que, chaque année, un ou plusieurs évaluateurs s'écartent systématiquement des autres évaluateurs dans la mesure où cet écart ne peut être attribué à une coïncidence. Suivant ces résultats nous avons effectué des analyses complémentaires, dites « comparaisons post-hoc », pour identifier le type de différence

⁵⁷ R.D. HAYS et D.A. REVIKI, "Reliability and validity (including responsiveness)", dans P. M. FAYERS et R. D. HAYS (red), *Assessing quality of life in clinical trials: Methods and practice*, NY, Oxford University Press, 2005; L.G. PORTNEY et M.P. WATKINS, *Foundations of clinical research: applications to practice. 3rd Edition*, New Jersey, Prentice Hall, 2009

⁵⁸ R.D. HAYS et D.A. REVIKI, "Reliability and validity (including responsiveness)", dans P. M. FAYERS et R. D. HAYS (red), *Assessing quality of life in clinical trials: Methods and practice*, NY, Oxford University Press, 2005; L.G. PORTNEY et M.P. WATKINS, *Foundations of clinical research: applications to practice. 3rd Edition*, New Jersey, Prentice Hall, 2009.

^{59 59} F.L. SCHMIDT et J.E. HUNTER, "The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings", *Psychological Bulletin* 1998, 262-274

⁶⁰ Ces facteurs sont les seuls recueillis et utilisables par le CSJ.

⁶¹ C'est à dire: le test du χ^2 pour deux variables nominales, le test de Student t ou l'analyse de variances pour une variable nominale et une variable continue, la corrélation singulière pour deux variables continues. Les analyses sont réalisées par le logiciel SPSS 24.0.

entre les examinateurs. Nous discernons deux types de différences. Premièrement, il existe des années où deux évaluateurs se positionnent à chaque extrême. Ces deux évaluateurs « extrêmes » évaluent les candidats différemment des autres évaluateurs et cela de façon statistiquement significative. Un évaluateur a tendance à attribuer un score nettement plus favorable aux candidats que les autres évaluateurs, pendant que l'autre évaluateur « extrême » a tendance à juger les candidats de façon nettement plus sévère. L'écart entre les scores moyens des deux évaluateurs extrêmes s'avère de deux à trois points. Deuxièmement, durant d'autres années on distingue deux groupes d'évaluateurs qui diffèrent l'un de l'autre de façon statistiquement significative. Un groupe a tendance à évaluer les candidats de façon plus favorable que l'autre groupe.

A titre d'illustration, le tableau 1 donne les scores moyens des huit évaluateurs pour l'épreuve écrite « rédaction d'un jugement » pour une certaine année. Le score moyen par évaluateur est la cote moyenne donnée par un évaluateur pour l'ensemble des candidats à cette épreuve. L'analyse de la variance décèle l'existence d'une différence significative entre les scores moyens des différents évaluateurs. Les comparaisons post-hoc montrent que cela s'explique avec les cotes des évaluateurs 1 et 5 qui s'écartent de façon significative de celles de tous les autres évaluateurs. On ne note pas de différences statistiquement significatives entre les autres évaluateurs.

EVALUATEUR	SCORE MOYEN
EVALUATEUR 1	11.72
EVALUATEUR 2	9.96
EVALUATEUR 3	10.00
EVALUATEUR 4	10.25
EVALUATEUR 5	8.45
EVALUATEUR 6	10.48
EVALUATEUR 7	10.33
EVALUATEUR 8	10.49

Tableau 1. Score moyen selon l'évaluateur ($n = 261$) pour la rédaction d'un jugement (*Fwelch* (7,2093) = 78,84, $p < 0.001$).

B. Concordance inter-évaluateur

Pour estimer le niveau de concordance entre les évaluateurs, une corrélation intra-classe (ICC) est calculée pour chaque épreuve écrite au cours de chaque année de sélection. Des corrélations inférieures au seuil de .80 sont considérées comme faibles pour des décisions concernant la sélection de personnel. Comme l'indique le tableau 2, le niveau de concordance entre les évaluateurs est excellent dans 84% des cas (27/32). Cependant, il existe des différences majeures dans le degré d'accord entre les évaluateurs selon les années. Certaines années, le degré d'accord est presque parfait (0.97 pour la résolution d'un cas pratique en 2012 et 0.96 pour la dissertation en 2007) ; d'autres années la corrélation intra-classe est nettement inférieure à .80 (0.63 pour la résolution d'un cas pratique en 2016, 0.69 pour la rédaction d'un jugement en 2002 et pour la dissertation en 2004). Nous

constatons que c'est surtout pour la correction de la dissertation qu'un niveau insuffisant d'accord entre les évaluateurs est rencontrée (3 fois sur 9 cas). Cela signifie que pour ces années-là, la variation dans les scores est due à des facteurs non pertinents (i.e. les évaluateurs) et non pas à la réelle variation dans la performance des candidats. Par rapport aux autres années, les notes des candidats pour ces années reflètent moins leur performance réelle et dépendent donc plus de la composition du jury d'examen.

TYPE D'ÉPREUVE	ANNÉE	COEFFICIENT DE CORRÉLATION INTRA-CLASSE
RÉDACTION D'UN JUGEMENT	2000	0.90
	2002	0.69
	2003	0.81
	2004	0.88
	2005	0.88
	2007	0.87
	2008	0.90
	2009	0.90
	2010	0.87
	2011	0.84
	2012	0.83
ANALYSE D'UN ARRÊT	2002	0.89
	2003	0.85
DISSERTATION	2000	0.77
	2001	0.85
	2004	0.69
	2005	0.75
	2007	0.96
	2008	0.87
	2009	0.93
	2011	0.88
	2012	0.89
RÉSOLUTION D'UN CAS	2012a	0.89
	2012b	0.97
	2013a	0.92
	2013b	0.85
	2014a	0.80
	2014b	0.89
	2015a	0.83
	2015b	0.89
	2016a	0.63
	2016b	0.91

Tableau 2. Coefficient de corrélation intra-classe par épreuve écrite par année

C. Nombre idéal d'évaluateurs

À notre connaissance, aucune motivation n'est formulée pour justifier le choix du nombre d'évaluateurs par épreuve. La justification d'un nombre déterminé repose sur la nécessité de garantir

une fidélité inter-évaluateur suffisante. Pour tester cela de manière empirique, nous utilisons la formule de Spearman Brown. Cette formule sert à estimer dans quelle mesure la fidélité d'un questionnaire s'accroît en fonction du nombre des items (fidélité *split half*). Récemment, cette formule est également utilisée pour déterminer le nombre d'évaluateurs nécessaire à l'obtention d'un degré suffisant de fidélité inter-évaluateur⁶². Grâce à cette formule nous avons pu calculer le coefficient de fidélité pour différents nombres d'évaluateurs (un évaluateur, deux évaluateurs, trois évaluateurs, quatre évaluateurs, etc.) et ceci pour chaque année et pour chacune des épreuves écrites (cf. les figures 3, 4, 5 et 6). Les résultats révèlent deux tendances. Tout d'abord, nous constatons que le coefficient de corrélation augmente avec le nombre d'évaluateurs. Cependant, cette augmentation n'est pas linéaire. Le coefficient de corrélation augmente de façon importante jusqu'à cinq évaluateurs. Par la suite l'augmentation est notablement moins prononcée, entraînant un coude dans la courbe. Le rendement par évaluateur supplémentaire au-delà de cinq évaluateurs est négligeable, particulièrement lorsque le degré d'accord entre évaluateurs est élevé. Ensuite, nous notons que la qualité des évaluateurs prime sur le nombre d'évaluateurs. Nous entendons par là que les années où le niveau d'accord est faible, une augmentation du nombre d'évaluateurs ne conduit pas nécessairement à une nette amélioration du coefficient de corrélation. Un plus grand nombre d'évaluateurs ne garantit donc pas un degré élevé d'accord.

⁶² H.C.W. DE VET, L.B. MOKKING, D.G. MOSMULLER et C.B. TERWEE, "Spearman-Brown prophecy formula and Cronbach's alpha : different faces of reliability and opportunities for new applications", *Journal of Clinical Epidemiology* 2017, 45-49

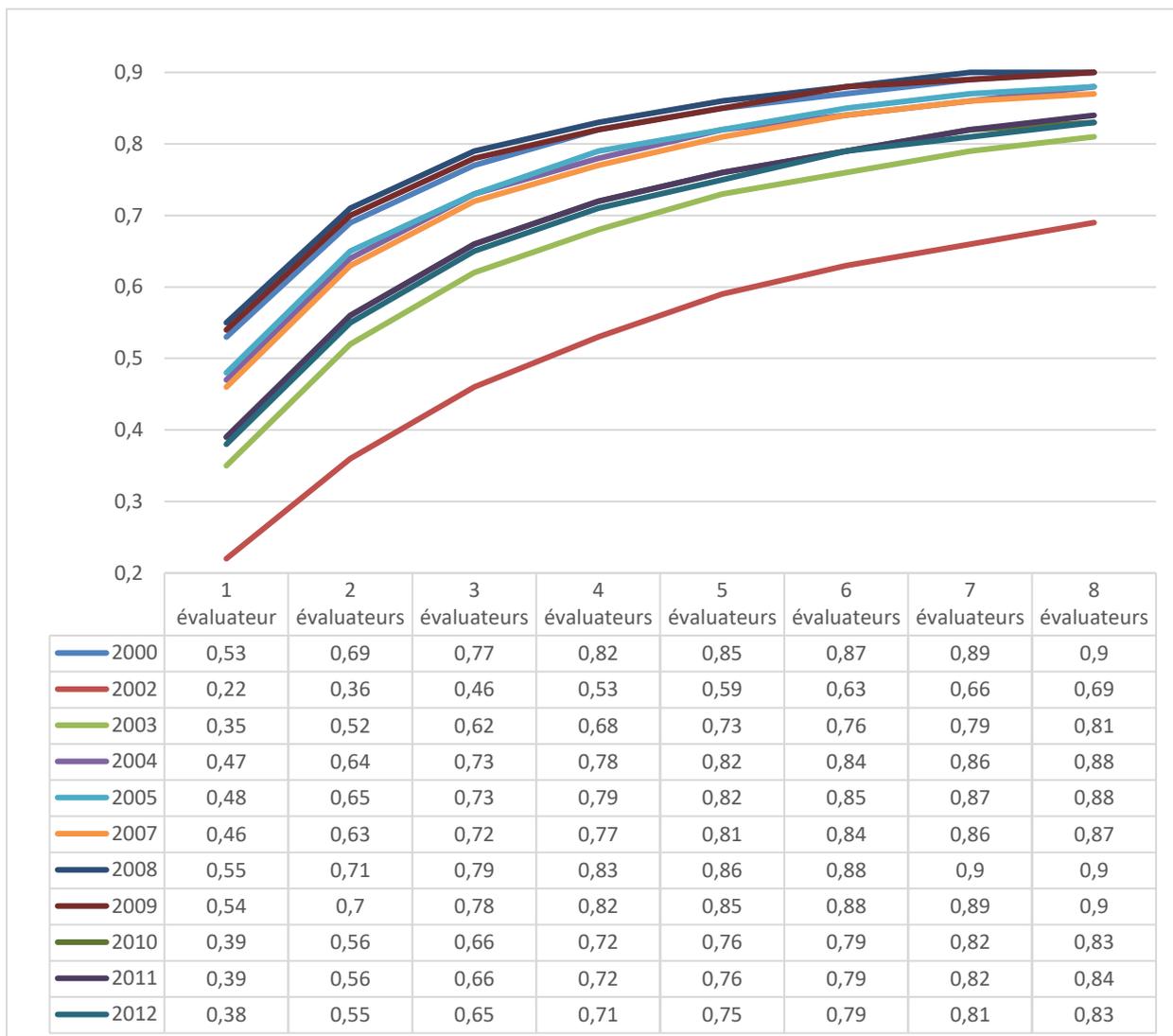


Fig. 3. Coefficient de corrélation selon le nombre d'évaluateurs pour la rédaction du jugement

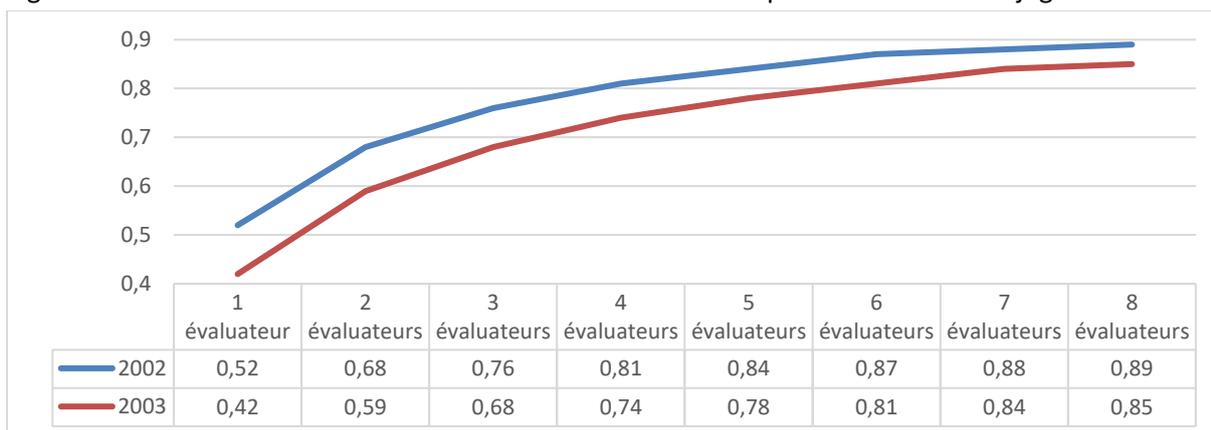


Fig. 4. Coefficient de corrélation selon le nombre d'évaluateurs pour l'analyse de l'arrêt

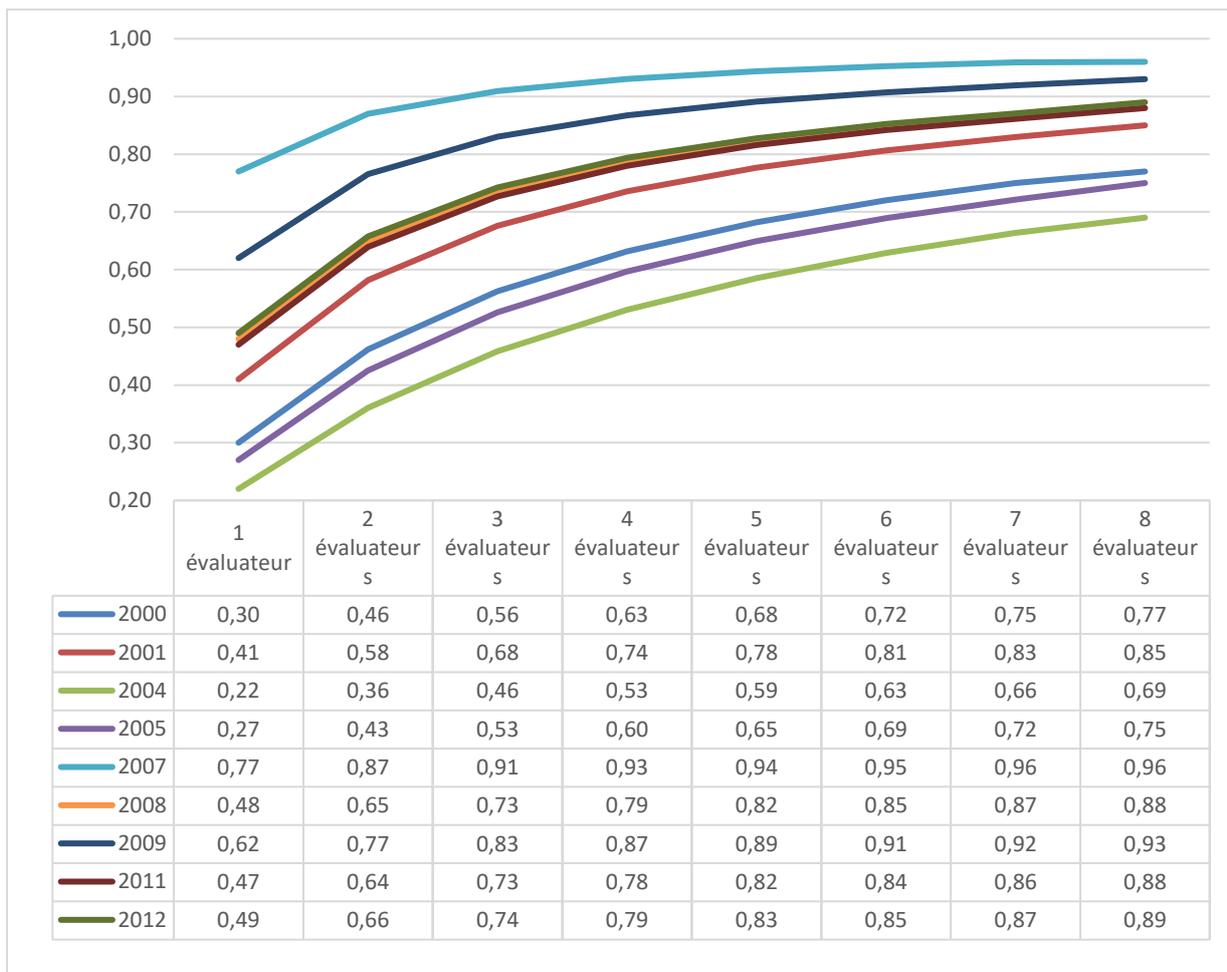


Fig. 5. Coefficient de corrélation selon le nombre d'évaluateurs pour la dissertation

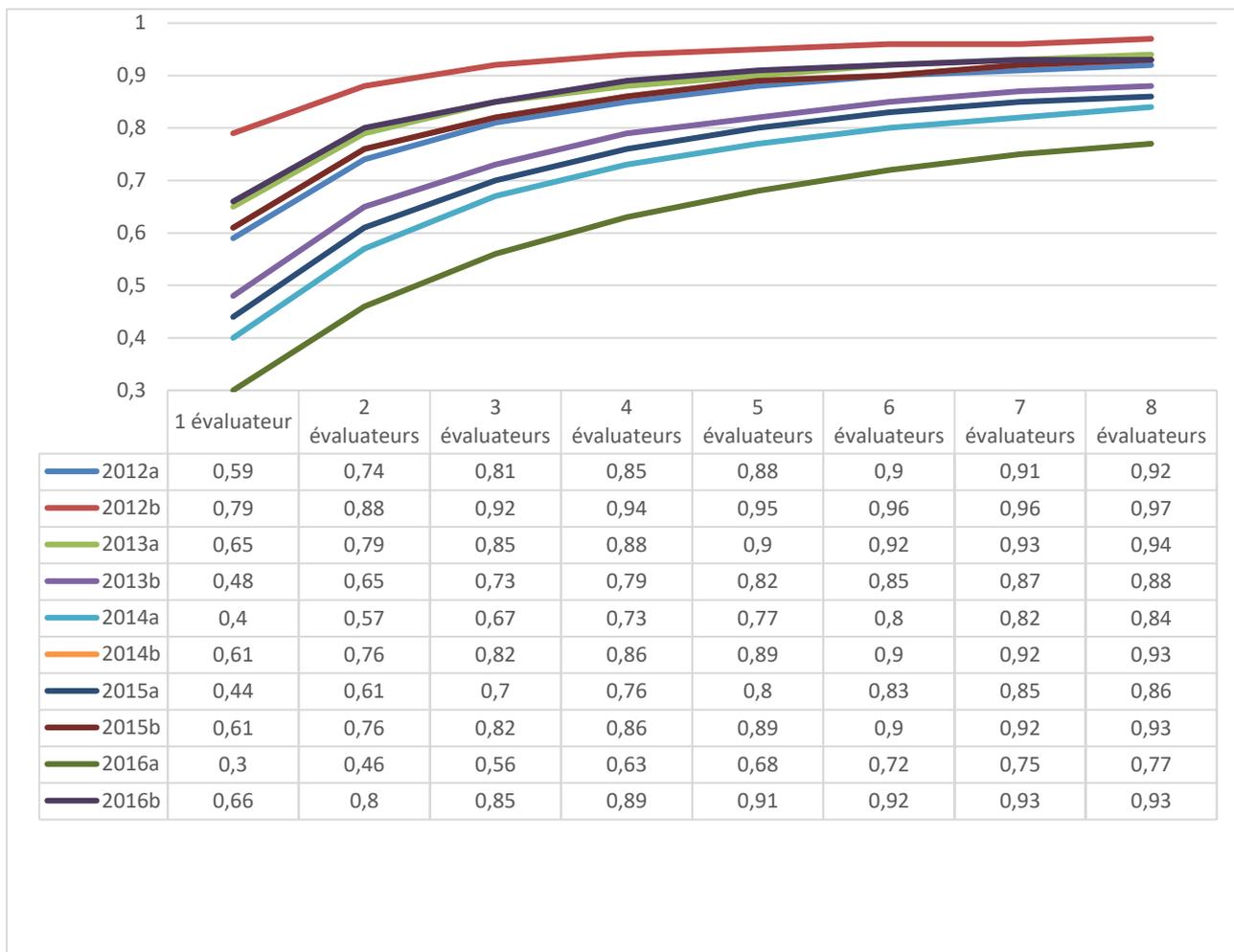


Fig. 6. Coefficient de corrélation selon le nombre d'évaluateurs pour la résolution du cas pratique

5. Résultats des analyses de la validité conceptuelle

La présentation des résultats des analyses de la validité conceptuelle est organisée en trois parties. Dans une première partie nous décrivons les candidats qui ont participé à l'examen d'aptitude professionnelle depuis l'instauration du CSJ en 2000. Les candidats sont décrits selon les facteurs étudiés dans cette contribution ; c'est-à-dire le sexe, l'âge, les grades académiques, l'université et le choix de la matière de droit. L'examen d'aptitude a été organisé à dix-neuf reprises (cet examen a eu lieu à deux reprises en 2012). Cependant les résultats des examens de l'année 2006 ne nous ont pas été fournis de même que ceux de la dernière année en 2017⁶³. Pour l'année 2010, une partie seulement des résultats des épreuves est disponible.

Ensuite, dans une deuxième partie nous abordons le taux de réussite à l'examen d'aptitude professionnelle. Le taux de réussite est présenté et analysé au cours des différentes phases de l'examen à savoir au terme de l'ensemble des épreuves (= le taux de réussite à l'examen), au terme

⁶³ Le CSJ nous a fourni de façon anonyme les données des candidats de l'examen d'aptitude professionnelle depuis son installation en 2000.

des épreuves écrites (= le taux de réussite à la totalité des épreuves écrites) et au terme de l'épreuve orale (= le taux de réussite à l'épreuve orale). Nous analysons en outre si les trois différentes périodes distinguées dans l'évolution de l'examen d'aptitude professionnelle (cf. fig. 1 et 2) influencent le taux de réussite.

Finalement, nous analysons l'influence des différents facteurs (sexe, âge, grades académiques, université et choix de la matière de droit) sur le taux de réussite. Nous maintenons une même structure dans l'analyse et la présentation des résultats que dans la deuxième partie : l'influence sur le taux de réussite à l'examen, à la totalité des épreuves écrites et à l'épreuve orale, et ceci selon les trois différentes périodes de l'évolution de l'examen d'aptitude professionnelle. Nous rappelons qu'uniquement les résultats statistiquement significatifs sont rapportés.

A. Description des candidats à l'examen d'aptitude professionnelle

Pour l'ensemble de la période étudiée, 1848 candidats ont passé l'examen ; parmi ceux-ci, 1059 l'ont présenté une seule fois (57,3%), 403, deux fois (21.8%), 194, trois fois (10.5%), 192, quatre fois ou plus (10.4%)⁶⁴. Cela conduit à globaliser 3429 participations.

Le nombre de participations par année a varié entre 168 en 2009 et 264 candidats en 2004 et la courbe de participation se dessine comme une double bosse de chameau (cf. fig. 7)

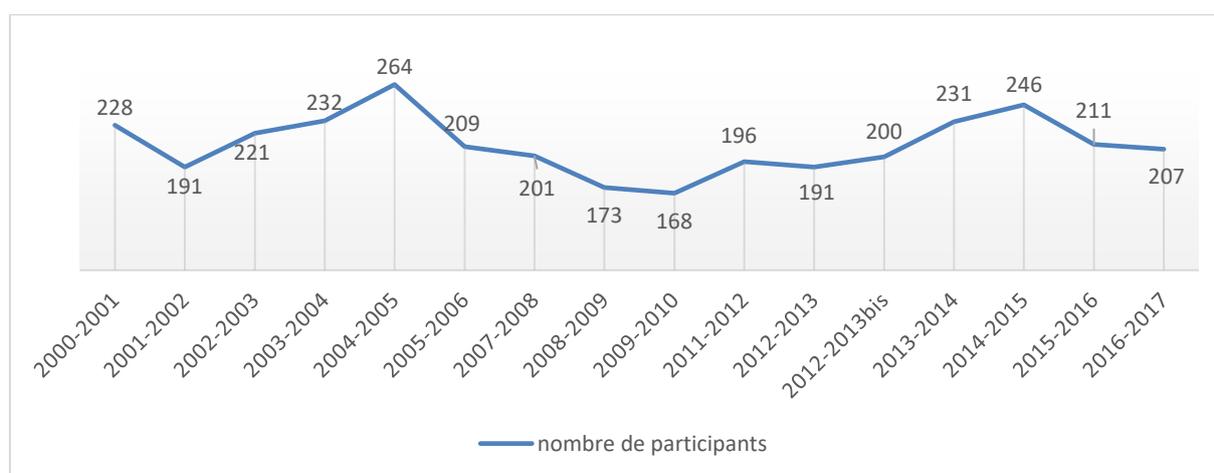


Fig. 7. Nombre de participants à l'examen d'aptitude au cours des années

Parmi les participants 58% sont des femmes, 42 % des hommes ; le ratio H/F des participants a évolué de façon significative dans le temps ($\chi^2(15) = 30.83, p = 0.009$) : comme en témoigne la fig. 8, le ratio initial de 1 s'est modifié progressivement pour passer à 2/3 au courant des dernières années.

⁶⁴ 96 quatre fois (5.2%), 38 cinq fois (2.1%), 31 six fois (1.7%), 14 sept fois (0.8%), 5 huit fois (0.3%), 2 neuf fois (0.1%), 3 dix fois (0.2%), 1 onze, douze et treize fois (0.1%). Le nombre de participations est seulement limité à cinq fois depuis 2016.

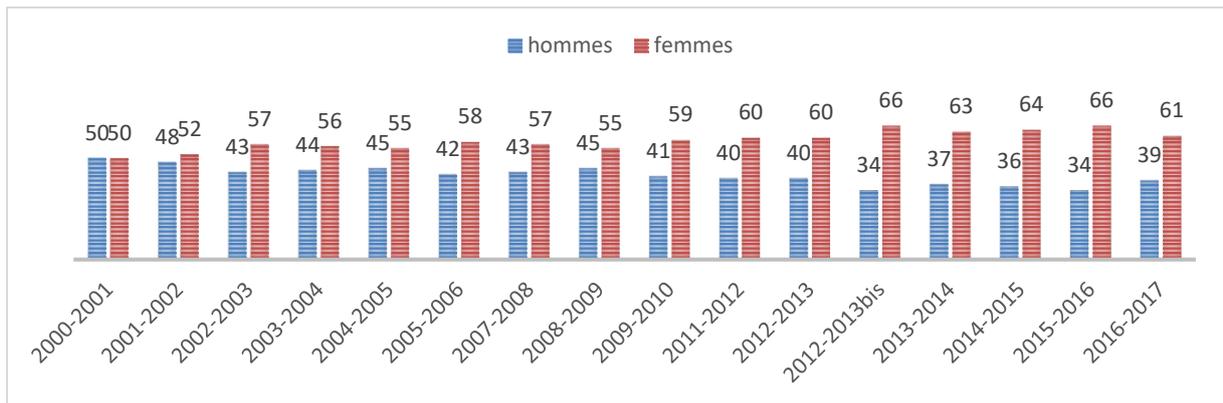


Fig. 8. Pourcentage des hommes et des femmes participant à l'examen d'aptitude au fil des années

L'âge moyen des participants à l'examen est de 35.6 ans ($sd = 6.9$ ans, écart : 22-64 ans) ; les hommes sont plus âgés que les femmes de manière significative ($H : 36.2, F : 35.2 ; t(3294) = 4.14, p < 0.001$). On note une grande disparité de l'âge moyen des hommes et des femmes durant notre étude (fig. 9 ; $F(16,3295) = 2.52, p < 0.001$).

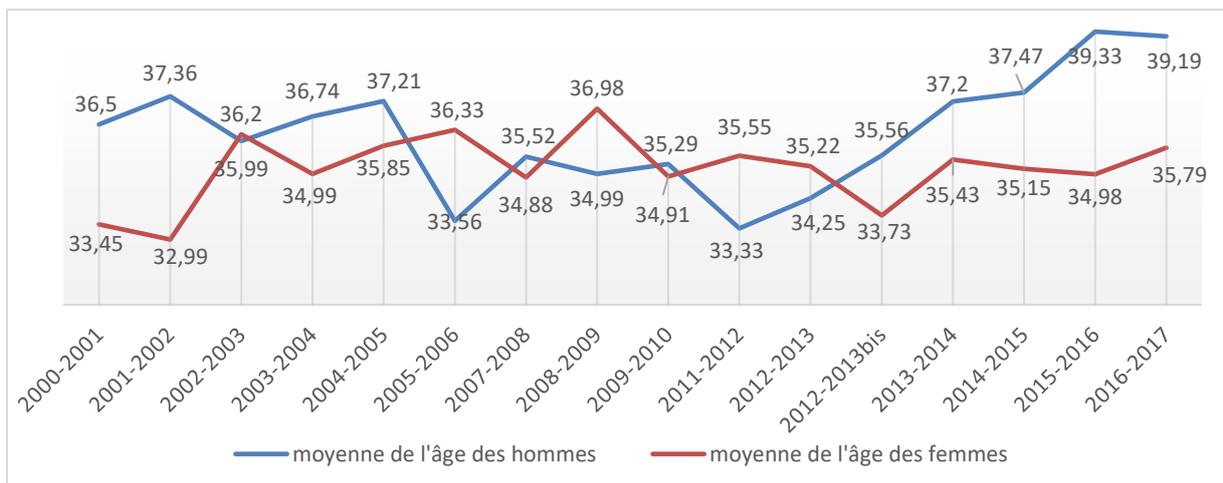


Fig. 9. Valeur moyenne de l'âge des hommes et des femmes participant à l'examen d'aptitude au fil des années

Les grades académiques des candidats se répartissent en 54% de satisfaction, 39 % de distinction et 7 % de grande distinction ou plus grande distinction. La répartition des grades au cours des années de l'étude ne s'est pas modifiée. On note cependant une différence de résultats entre les hommes et les femmes ($\chi^2(3) = 9.79, p = 0.020$); ces dernières ont de meilleurs résultats que les hommes (fig. 10) et cette tendance s'accroît au cours de la période étudiée.

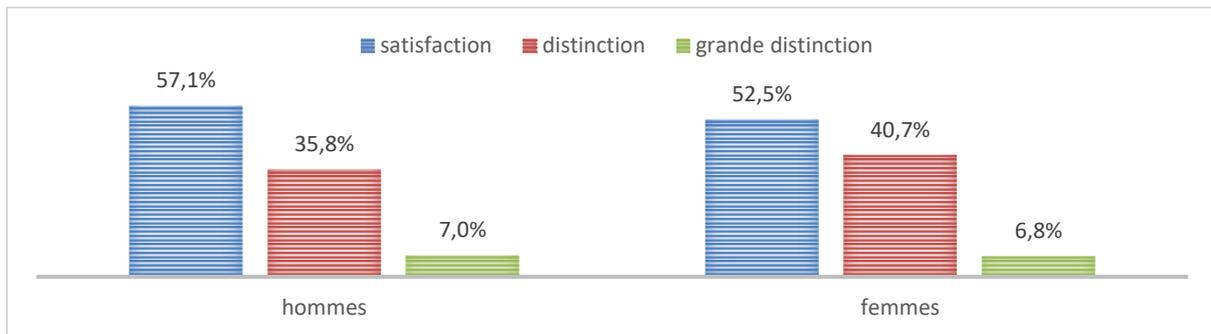


Fig. 10. Répartition des grades académiques pour les candidats à l'examen d'aptitude selon leur sexe

Les quatre universités néerlandophones ont diplômé les candidats suivant la répartition : KU Leuven (46%), UGent (26%), UA (17%), VUB (11%). Il apparaît que le rapport hommes-femmes varie selon les universités (cf. VUB et KU Leuven; fig. 11 ; $\chi^2(2) = 8.16, p = 0.017$).

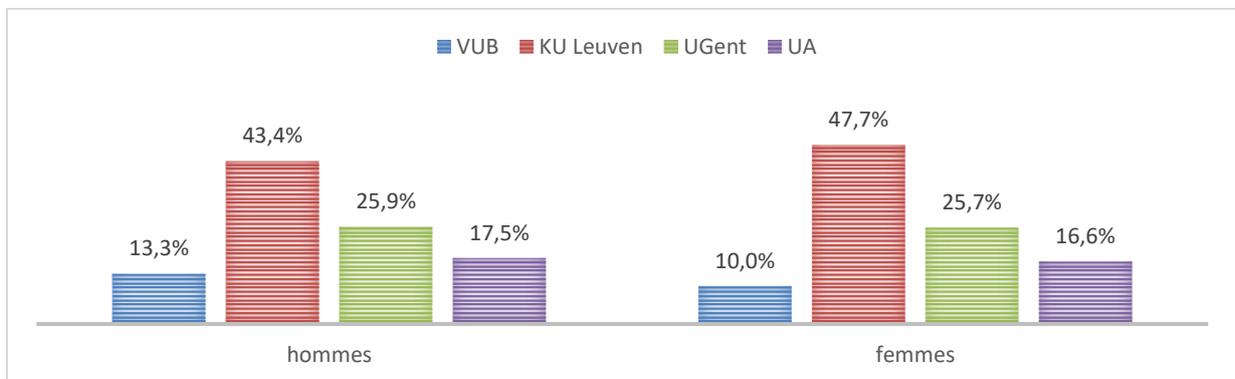


Fig. 11. Répartition des universités pour les candidats à l'examen d'aptitude selon leur sexe

La répartition des grades par université montre une nette prépondérance des grades pour l'UGent ($\chi^2(6) = 129.67, p < 0.000$) où le ratio distinction ou plus/satisfaction voisine le 6/4. Dans les trois autres universités ce ratio est inversé (fig. 12).

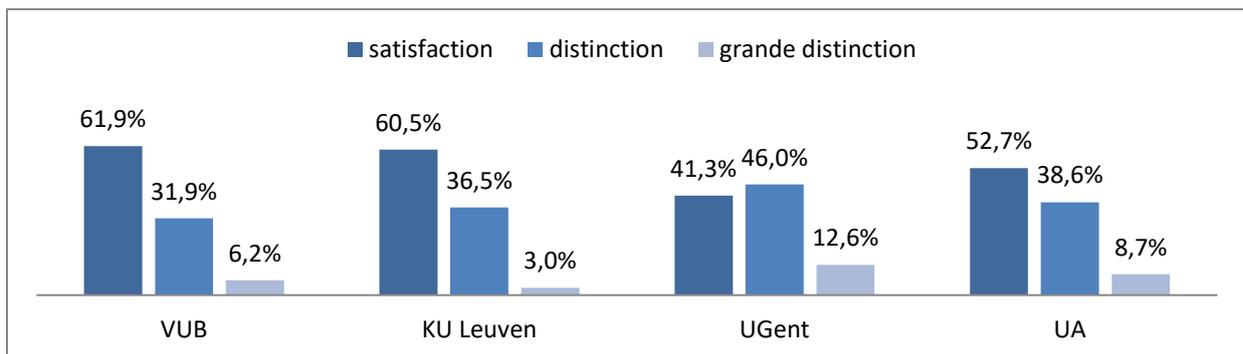


Fig. 12. Répartition des grades pour les candidats à l'examen d'aptitude selon l'université

La matière de droit pour l'examen, laissée au choix des candidats, se répartit en droit civil (59%), droit pénal (30%), droit social (11%). On remarque que cette répartition se modifie dans le temps, (fig. 13 ; $\chi^2(30) = 75.38, p < 0.000$). Notamment en 2012, lorsque l'examen a été organisé à deux reprises, la proportion des candidats choisissant le droit pénal et le droit civil est presque égale. En 2001, il n'y avait pas d'option de droit social.

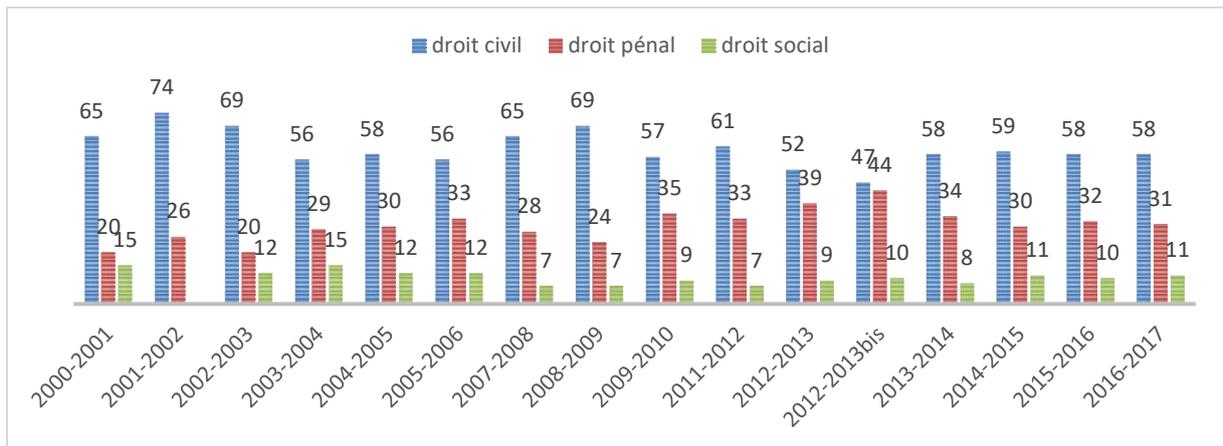


Fig. 13. Répartition de la matière de droit pour les candidats à l'examen d'aptitude au fil des années

De plus, le droit pénal est choisi de façon significative par les femmes ($\chi^2(2) = 6.82, p = 0.033$), par les candidats plus jeunes ($F(2,3036) = 26.85, p < 0.000$) et ceux diplômés sans grade académique ($\chi^2(4) = 11.75, p = 0.019$).

B. Taux de réussite à l'examen

i. Réussite à la totalité des épreuves

Le taux de réussite de la totalité des épreuves s'élève à 10% pour l'ensemble des années ; cependant on note des différences importantes au cours des années (de 6 à 14.9% ; $\chi^2(15) = 27.75, p = 0.023$) (fig. 14).

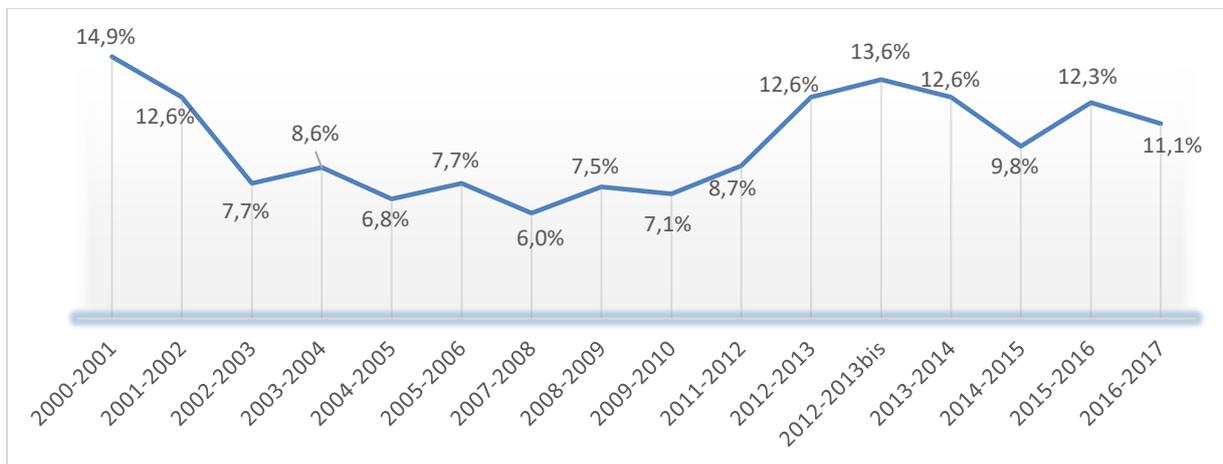


Fig. 14. Taux de réussite à l'examen d'aptitude au fil des années

Une analyse du taux de réussite à l'examen selon les trois périodes de l'évolution du processus révèle que ce processus influence nettement le taux de réussite : à l'instauration de la deuxième période (en 2002), qui se caractérise par un système d'épreuves écrites en cascade, le taux de réussite est passé de 13.8 à 8.8% ; par contre, le taux de réussite est remonté à 11.8% lors de la troisième période (en 2012) quand cette double épreuve en cascade est remplacée par une épreuve écrite unique ($\chi^2(2) = 18.75, p < 0.000$).

ii. Réussite aux épreuves écrites

Le taux de réussite des épreuves écrites s'élève à 20.1% pour l'ensemble des années et suit la même tendance que celle observée pour l'ensemble de l'examen : importante variabilité au cours des années (11.4 à 32.5% ; $\chi^2(16) = 188.76, p < 0.000$) (fig. 15) ; et nette différence selon les trois périodes (32.5% pour la première période, à 15.9% pour la deuxième période et 21.1% pour la troisième période ; $\chi^2(2) = 41.53, p < 0.000$).

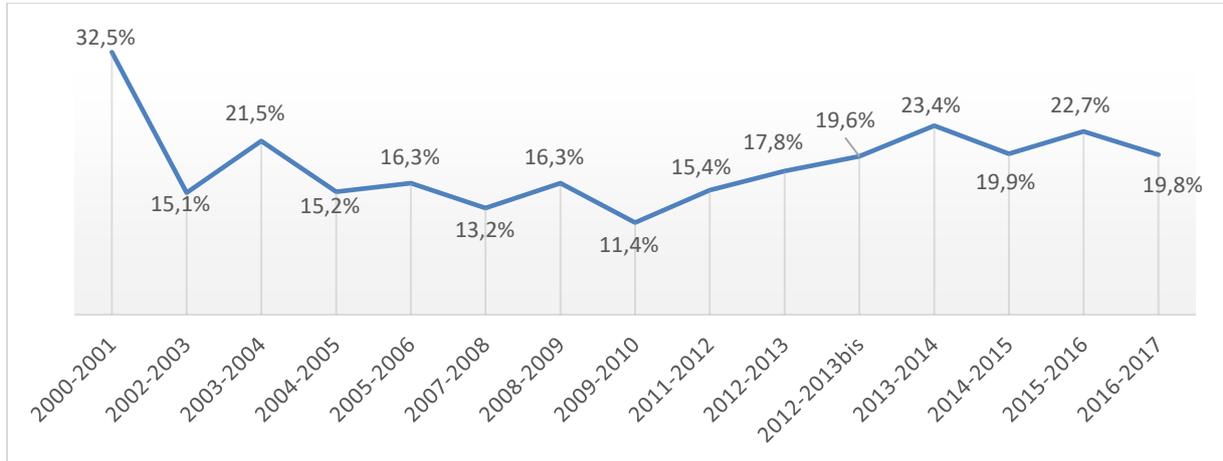


Fig. 15. Taux de réussite aux épreuves écrites de l'examen d'aptitude au fil des années

La deuxième période consiste en l'organisation de deux épreuves écrites éliminatoires. Nous analysons les résultats obtenus à chaque étape de l'examen écrit. Le taux moyen de réussite pour l'ensemble des premières épreuves s'élève à 37.6% ; cependant le taux de réussite varie grandement en fonction du contenu de cette épreuve : 60.3% pour l'analyse de l'arrêt de Cassation, 40.5% pour le QCM et dissertation, 34.5% pour la rédaction d'un jugement et 31% pour la dissertation ($\chi^2(4) = 70.28, p < 0.000$). Quant à la seconde épreuve écrite, le taux moyen de réussite est de 43% ; à nouveau de grandes différences existent entre le contenu des épreuves : 51.4% pour l'analyse de l'arrêt, 48.5 pour la dissertation et 40.1 pour le jugement ($\chi^2(2) = 6.75, p = 0.034$). Nous remarquons que la rédaction du jugement conduit à des résultats médiocres et que l'analyse de l'arrêt de cassation obtient le meilleur score (fig. 16).

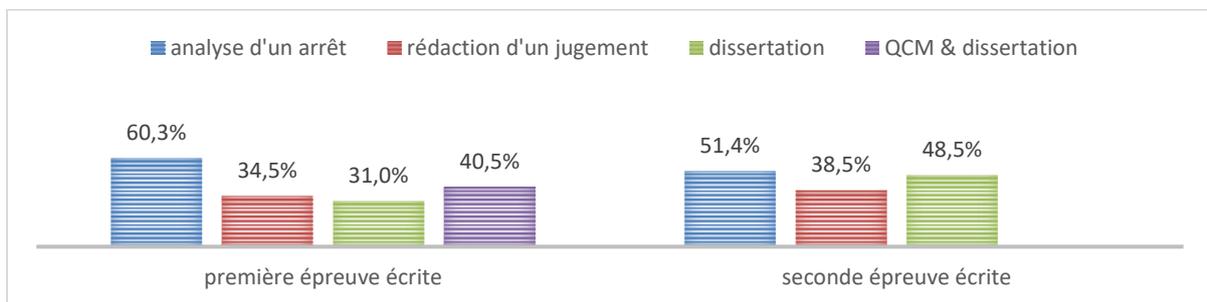


Fig. 16. Taux de réussite à la première et à la deuxième épreuve écrite selon le contenu de l'épreuve

iii. Réussite à l'épreuve orale

L'épreuve orale est éliminatoire pour 48.6% de l'ensemble des lauréats des épreuves écrites et cette valeur reste stable pour l'ensemble des années de l'examen. Le processus d'organisation des épreuves écrites n'a pas influencé le taux de réussite de l'épreuve orale.

C. Facteurs influençant le taux de réussite

i. Sexe

Sur l'ensemble des années le taux de réussite, que ce soit pour l'ensemble de l'examen, au terme des épreuves écrites ou pour l'épreuve orale, est identique pour les hommes et les femmes à savoir 10.5 versus 10% pour l'ensemble de l'examen, 20.1 versus 19.2% pour la totalité des épreuves écrites, 52.1 versus 51.0% pour l'épreuve orale, et cela quel que soit le processus d'examen.

Des différences apparaissent cependant à l'analyse du contenu des épreuves écrites : au QCM le taux de réussite est de 58.1 pour les hommes et de 28.8 % pour les femmes ($\chi^2(1) = 47.37, p < 0.000$); cette différence de résultat se répercute sur le résultat final de l'examen (13.5 contre 6.9% ; $\chi^2(1) = 6.67, p = .010$). Il semble également, par ailleurs, que la séquence des épreuves joue un rôle sur le taux de réussite des hommes et des femmes ; en effet, la séquence dissertation suivie de la rédaction d'un jugement a un impact favorable pour les femmes sur le résultat final (9.6 contre 3.0% ; $\chi^2(1) = 6.45, p = 0.011$) (fig. 17)⁶⁵.

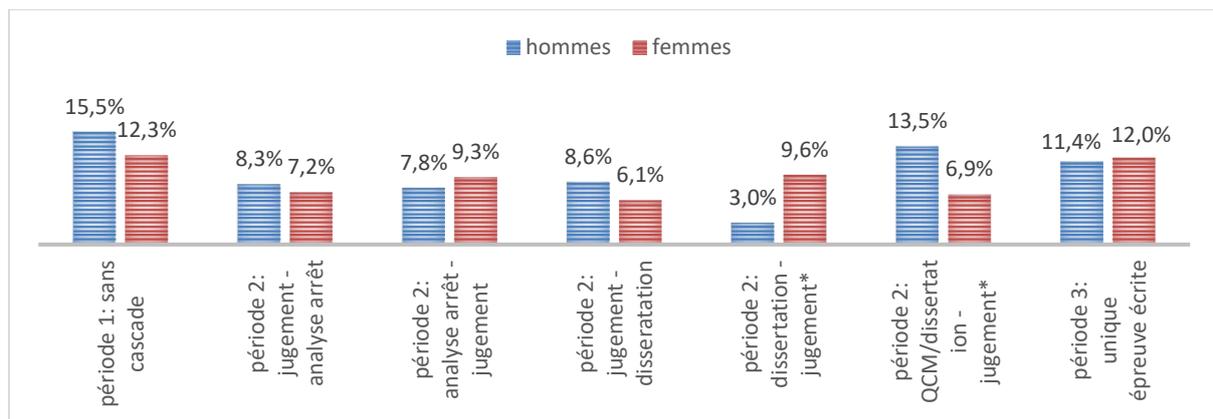


Fig. 17. Taux de réussite à l'examen d'aptitude selon le sexe des candidats et selon le contenu de l'examen au cours des trois périodes

ii. Âge

Sur l'ensemble des années, les lauréats de l'examen sont plus jeunes (34.8 années) que les échoués (35.7 années) ($t(3131) = -2.64, p = 0.005$). Comme le montre le tableau 3, cette différence apparaît uniquement dans la résolution d'un cas (période 3) et lors de la séquence dissertation – jugement et jugement – analyse d'un arrêt.

⁶⁵ Dans la figure un astérisque indique les différences statistiquement significatives

Type d'examen		Moyenne de l'âge	Test Student t	Valeur p
Période 1 : cascade			0.91	0.365
	Lauréats	35.79		
	Non-lauréats	34.83		
Période 2 : jugement – analyse d'un arrêt			-2.81	0.005
	Lauréats	31.53		
	Non-lauréats	36.49		
Période 2 : analyse d'un arrêt – jugement			0.69	0.490
	Lauréats	36.80		
	Non-lauréats	35.67		
Période 2 : jugement – dissertation			-0.53	0.594
	Lauréats	35.18		
	Non-lauréats	35.89		
Période 2 : dissertation – jugement			-4.28	0.000
	Lauréats	31.92		
	Non-lauréats	35.85		
Période 2 : QCM & dissertation – jugement			1.42	0.157
	Lauréats	35.81		
	Non-lauréats	34.70		
Période 3 : résolution d'un cas			-3.13	0.009
	Lauréats	34.49		
	Non-lauréats	36.31		

Tableau 3. Moyenne de l'âge des candidats selon le contenu de l'examen pour les lauréats et les non-lauréats de l'examen

iii. Grade académique

Sur l'ensemble des années, le taux de réussite pour l'examen et pour la totalité des épreuves écrites dépend de façon significative du grade académique du candidat et se répartit pour l'examen : 7.1% pour la satisfaction, 12.7% pour la distinction et 18.8% pour la grande distinction ou plus ($\chi^2(2) = 43.98$, $p < 0.000$) ; et pour les épreuves écrites : 14.7% pour la satisfaction, 23.7% pour la distinction et 26.7% pour la grande distinction ou plus ($\chi^2(2) = 40.33$, $p < 0.000$).

En analysant les différentes périodes de l'examen et le contenu des épreuves écrites, on remarque que certaines épreuves écrites sont associées de façon statistiquement significative au taux de réussite finale des candidats porteurs d'un grade : à savoir la dissertation suivie d'un jugement ($\chi^2(2) = 9.86$, $p = 0.007$) et le QCM ($\chi^2(2) = 24.74$, $p < 0.000$) (lors de la deuxième période), et la résolution du cas pratique (lors de la troisième période ; $\chi^2(2) = 15.93$, $p < 0.000$) (fig. 18)⁶⁶. L'analyse des épreuves individuelles de la deuxième période démontre que c'est la dissertation et le QCM et non pas le jugement qui provoquent ces différences selon les grades du candidat.

⁶⁶ Dans la figure un astérisque indique les différences statistiquement significatives

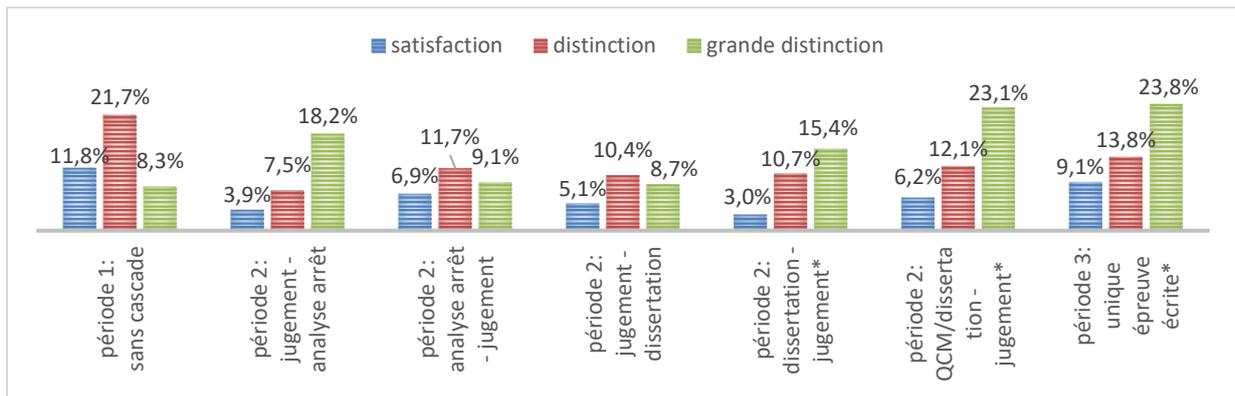


Fig. 18. Taux de réussite à l'examen d'aptitude selon le grade académique des candidats et selon le contenu de l'examen durant les trois périodes

Il en est de même pour le type d'examen oral : uniquement lors de la seconde période, c'est-à-dire la discussion à partir d'un cas, le taux de réussite varie statistiquement selon le grade académique (fig. 19 ; $\chi^2(2) = 11.96, p = 0.003$)⁶⁷.

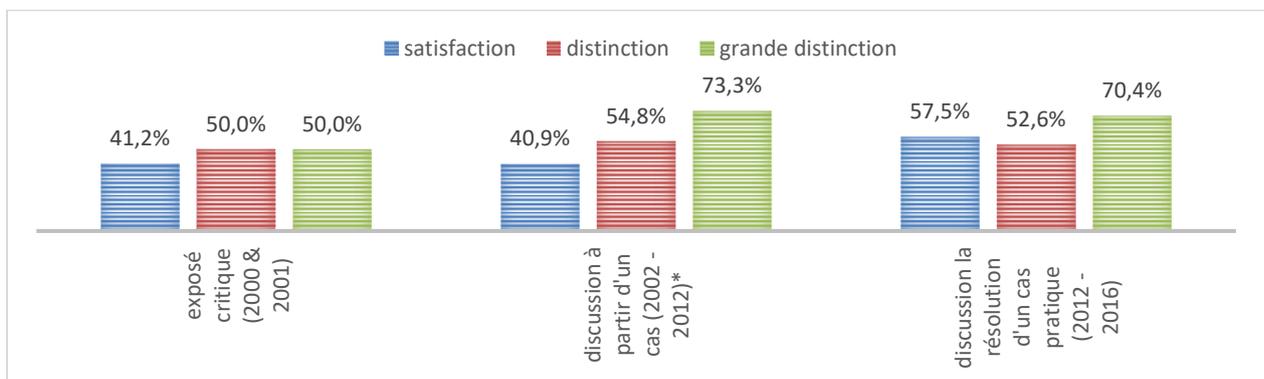


Fig. 19. Taux de réussite à l'épreuve orale de l'examen d'aptitude selon le grade académique des candidats et selon le contenu de l'épreuve orale

iv. Université

Le taux de réussite à l'examen varie en fonction de l'université du candidat pour l'ensemble des années étudiées: VUB: 4.6%, KU Leuven: 11.9%, UGent 10.5% et UA 9.9% ($\chi^2(3) = 15.30, p = 0.002$). Cependant, l'analyse selon les différentes périodes dans le processus d'examen, décèle que cette différence ne se marque de façon statistiquement significative que lors de la deuxième ($\chi^2(3) = 8.53, p = 0.036$) et la troisième période ($\chi^2(3) = 8.09, p = 0.044$)(fig. 20). Une même tendance est notée pour la réussite aux épreuves écrites.

⁶⁷ Dans la figure un astérisque indique les différences statistiquement significatives

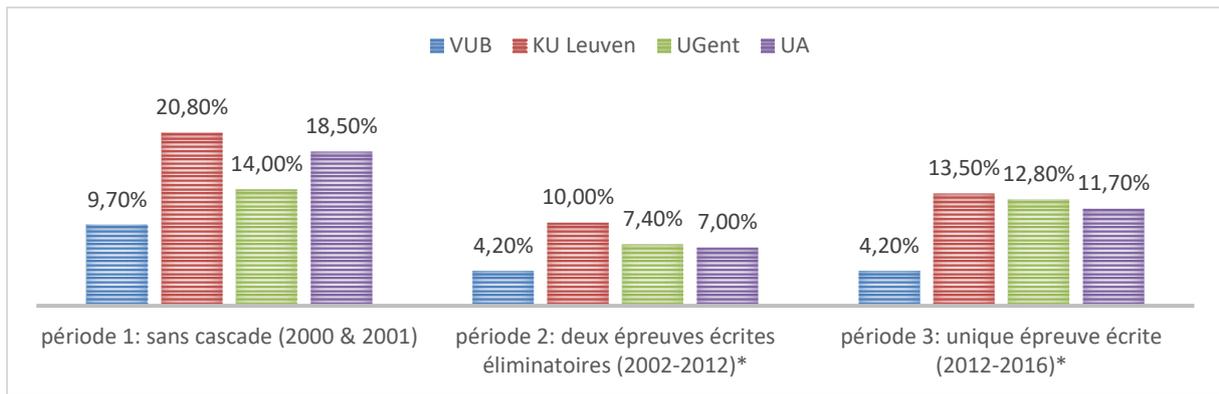


Fig. 20. Taux de réussite à l'examen d'aptitude selon l'université des candidats et selon le processus de l'examen

Ensuite, nous effectuons une analyse du taux de réussite des différentes épreuves écrites lors de la deuxième période. Cette analyse démontre que les candidats issus de la VUB sont plus faibles quel que soit le contenu de l'épreuve. Mais c'est uniquement dans l'épreuve QCM/dissertation que cette différence est statistiquement significative (fig. 21).

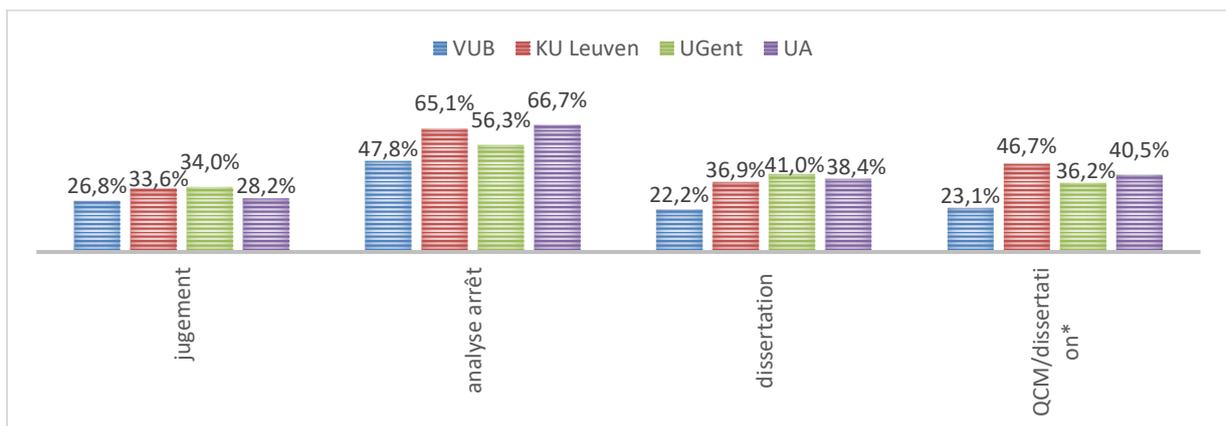


Fig. 21. Taux de réussite à la première épreuve écrite selon l'université des candidats et selon le contenu de cette épreuve (période 2)

Aucune différence dans le taux de réussite n'est remarquée à la deuxième épreuve écrite, ni à la réussite à l'épreuve orale selon l'université du candidat.

v. Matière de droit

Le taux de réussite pour l'examen ne diffère pas de façon statistiquement significative selon le choix de la matière de droit : droit civil 9.8%, droit pénal 11.0% et droit social 14.2%. Il en est de même pour la réussite aux épreuves écrites : droit civil 19.7%, droit pénal 20.2% et droit social 25.0%.

Cependant, l'analyse du taux de réussite des épreuves écrites séparées révèle une différence significative pour la rédaction du jugement. Les candidats qui rédigent un jugement dans le domaine du droit social réussissent plus souvent que les autres candidats (droit social : 40.7%, droit pénal : 33.9% et droit civil : 27.9% ; $\chi^2(2) = 6.17, p = 0.046$) (fig. 22). Notons que seuls le jugement et l'analyse de l'arrêt étaient différents pour la matière de droit.

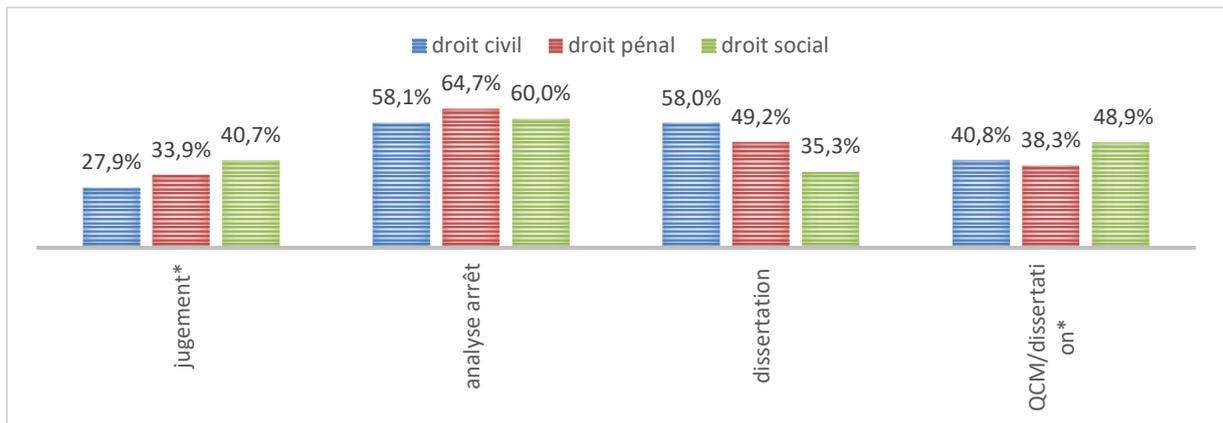


Fig. 22. Taux de réussite à la première épreuve écrite selon la matière de droit et selon le contenu de cette épreuve (période 2)

Cette différence ne se retrouve cependant pas pour la deuxième épreuve écrite, ni pour l'épreuve orale.

6. Conclusions

Notre conclusion se décompose en deux parties au cours desquelles nous abordons les résultats les plus saillants des analyses de la fidélité inter-évaluateur, d'une part, et de la validité conceptuelle d'autre part. Nous pensons pouvoir évoquer des pistes d'amélioration pour remédier aux difficultés révélées par ces analyses.

A. Conclusions concernant la fidélité inter-évaluateur

Concernant la fidélité inter-évaluateur nous retenons deux conclusions principales. Tout d'abord, nous constatons que des facteurs non-pertinents, portant sur le degré de sévérité de l'évaluateur, influencent la cote attribuée aux candidats. En effet, chaque année et pour chaque type d'épreuve écrite, les cotes d'un ou de plusieurs examinateurs s'écartent de manière significative des autres examinateurs. Dans la mesure où cela peut avoir des conséquences préjudiciables sur la qualité de la sélection (risque de sous-estimation ou de surestimation des compétences des candidats), une stratégie de gestion de ces scores déviants doit être élaborée. A cette fin, après chaque examen il s'impose de rechercher de manière systématique les examinateurs déviants ainsi que de déterminer le seuil d'écart maximal. Après tout, un certain degré d'écart est très habituel lorsque plusieurs personnes doivent évaluer le même test. Un outil permettant de déterminer ce seuil d'écart est la signification statistique. Cette mesure de signification est courante dans les recherches empiriques et a également servi à interpréter les résultats dans notre contribution. Ensuite, il est impératif de minimiser l'impact des évaluateurs déviants sur les résultats des participants. Cela peut conduire, soit à l'exclusion des résultats de cet évaluateur lors du calcul des scores finaux des candidats, soit à

l'ajustement systématique des scores attribués par cet évaluateur, soit à l'application des mécanismes de correction statistique⁶⁸.

Une seconde conclusion met en évidence la nécessité d'optimiser le degré de concordance entre les différents évaluateurs et d'améliorer ainsi l'objectivité de la méthode de sélection. Nous notons que le degré d'accord minimal de 0,80, validé dans la littérature, n'est pas atteint toutes les années. Il en ressort que, ces années-là, la composition du jury a pu influencer sur les résultats des participants. De plus, nos résultats ne soutiennent pas la nécessité d'un jury de huit membres. Un degré d'accord suffisant est déjà atteint avec quatre ou cinq évaluateurs, pour autant que la qualité de ces évaluateurs soit garantie. Il est donc plus opportun d'optimiser la qualité des évaluations que d'accroître le nombre de membres du jury. Plusieurs recommandations d'optimisation sont formulées dans la littérature pour améliorer la concordance inter-évaluateur : (1) prévoir une formation continue et approfondie sur les facteurs à même de fausser une évaluation objective (*rater error training*)⁶⁹ de même qu'une formation spécifique sur le contenu de l'évaluation et les compétences évaluées (*frame of reference training*)⁷⁰ ; (2) prévoir un canevas de correction détaillé qui inclut non seulement toutes les réponses correctes, mais également les réponses incomplètes et les réponses erronées, et précisant les scores adaptés au type de réponse⁷¹; (3) élaborer une politique de sélection des évaluateurs selon des critères de sélection suivants⁷² : concordance des valeurs personnelles avec les fondements théoriques de la méthode de sélection⁷³ ; la motivation des évaluateurs ; l'acceptation de l'objectif du système d'évaluation⁷⁴ ; la vision commune sur le «magistrat performant»⁷⁵.

B. Conclusions concernant la validité conceptuelle

Avant d'aborder les conclusions concernant la validité conceptuelle, nous soulignons le taux particulièrement faible de réussite à l'examen qui se limite à un candidat sur dix, alors que chaque candidat est porteur d'un diplôme de licencié/master en droit, et a acquis une expérience juridique depuis plusieurs années. Nous admettons qu'il est indispensable de prévoir une sélection sévère en raison du rôle notable et essentiel du juge dans notre société. En revanche, ce taux de réussite soulève des questions concernant la qualité de l'enseignement juridique. La formation juridique acquise avec un diplôme de master suffit-elle réellement pour réussir l'examen ou faut-il repenser non seulement

⁶⁸ W.T. HYOT, "Rater bias in psychological research: When is it a problem and what can we do about it?", *Psychological Methods* 2000,64-86.

⁶⁹ D.J. WOEHR et A.I. HUFFCUT, "Rater training for performance appraisal: A quantitative review", *Journal of Occupational and Organizational Psychology* 1994, 189-205; F. LIEVENS, "Assessor training strategies and their effects on accuracy, interrater reliability, and discriminant validity", *Journal of Applied Psychology* 2000, 255-264; N.M.A. HAUENSTEIN & M.E. McCUSKER, "Rater training: Understanding effects of training content, practice ratings, and feedback", *International Journal of Selection and Assessment* 2017, 253-266

⁷⁰ S.G. ROCH, D.J. WOEHR, V. MISHRA et U KIESZCZYNSKA, "Rater training revisited: An updated meta-analytic review of frame-of-reference training", *Journal of Occupational and Organizational Psychology* 2012, 370-395

⁷¹ H.J.M. VAN BERKEL et A.E. BAK, *Toetsen in het Hoger Onderwijs*, Houten, Bohn Stafleu van Loghum, 2006

⁷² R.L. JOHNSON, J.A. PENNY et B. GORDON, *Assessing performance*, New York, NY, The Guilford Press, 2008

⁷³ A.E. Henry, K.J. Grimm, et R.C. Pianta, "Rater calibration when observational assessment occurs at large-scale: Degree of calibration and characteristics of raters associated with calibration.", thèse de doctorat, University of Virginia, 2010, <http://proquest.umi.com/pqdlink?Ver=1&Exp=08-31-2016&FMT=7&DID=2396570251&RQT=309&attempt=1&cfc=1>

⁷⁴ A. TZINER, K.R. MURPHY et J.N. CLEVELAND, "Contextual and rater factors affecting rating behavior", *Group & Organization Management* 2005, 89-98

⁷⁵ C. DANIELSON, "Evaluations that helps teachers learn", *Educational research* 2011, 35-40

l'enseignement actuel mais également les critères d'admission à l'examen ? Ce taux faible suggère-t-il la nécessité d'organiser une formation complémentaire de magistrat ?

Nos constatations révèlent que l'influence des facteurs non pertinents sur le taux de réussite diffère considérablement selon le contenu de l'examen. Par exemple, les femmes, quel que soit leur grade académique, sont nettement moins performantes devant un questionnaire à choix multiple. Cette découverte est corroborée par d'autres recherches scientifiques : les hommes sont plus performants aux QCM, indépendamment de la connaissance des participants sur le matériel mis en cause⁷⁶. Les QCM, plus spécifiquement « la correction pour devinette », mesurent outre des connaissances du candidat, également le comportement à risque des participants⁷⁷, généralement plus élevé chez les hommes⁷⁸. Un questionnaire à choix multiple n'est donc pas une méthode neutre pour mesurer des connaissances. Il est bon de noter que le contenu de l'examen d'aptitude professionnelle actuel – sans QCM - ne démontre pas ce biais du sexe.

Le taux de réussite à l'examen actuel est en revanche influencé par d'autres facteurs : l'université et le grade académique. Les participants de certaines universités (notamment la VUB) sont désavantagés par rapport aux autres participants. Bien que cette conclusion soit sans aucun doute riche de renseignements, l'analyse des résultats ne permet pas d'expliquer ces différences de manière concluante. S'agit-il d'une différence réelle dévoilée par l'examen ? C'est-à-dire, est-ce que l'examen met en évidence l'insuffisance de connaissances/compétences chez des juristes issus de certaines universités ? Ou bien la méthode de sélection concorde-t-elle mieux avec les connaissances acquises dans certaines universités ? Dans ce cas, la méthode de sélection actuelle n'est pas neutre. On peut éventuellement avancer que l'épreuve actuelle, qui se base sur l'*affective legal analysis*⁷⁹, a été développée par un membre du corps académique de la KU Leuven qui enseigne ce type de raisonnement aux futurs juristes. Par le fait même cela pourrait en partie expliquer les résultats favorables des candidats émanant de cette université. Une autre hypothèse se porte sur la composition du jury. Le jury qui rédige et qui corrige les épreuves est-il réparti équitablement entre les universités ? Chaque faculté a probablement ses caractéristiques et sa manière de transmettre le savoir. On pourrait alors supposer qu'un déséquilibre dans la répartition du jury pourrait influencer la rédaction et l'évaluation de l'épreuve suite aux différences entre universités, ce qui pourrait conduire à avantager involontairement certains candidats.

En outre, il se peut que le contenu des programmes et la politique d'évaluation des étudiants varient fondamentalement selon les différentes facultés de droit. Seule une analyse détaillée du programme des différents masters en droit, le taux de réussite des différentes facultés et de la personnalité des

⁷⁶ G. ANDERSON, D. BENJAMIN et M.A. FUSS, "The determinants of success in university introductory economics courses", *Journal of Economic Education* 1994, 99-119; G. STOBART, J. ELWOOD et M. QUINLAN, "Gender bias in examinations: how equal are the opportunities?", *British Educational Research Journal* 1992, 261-276; S.M. LINDBERG, J.S. HYDE et J.L. PETERSEN, "New trends in gender and mathematic performance: a meta-analysis", *Psychological Bulletin* 2010, 1123-1135; G. GERHARD et V. WAGNER, "Shying Away from demanding Tasks? Experimental Evidence on gender differences in answering multiple-choice questions", *Economics of Education Review* 2017, 43-62

⁷⁷ W.E. BECKER et C. JOHNSTON, "The relationship between multiple choice and essay response questions in assessing economics understanding", *Economic Record* 1999, 348-357

⁷⁸ J.P. BYRNES, D.C. MILLER et W.D. SCHAFER, "Gender differences in risk taking: a meta-analysis", *Psychological Bulletin* 1999, 367-383

⁷⁹ F. FLEERACKERS, *Affective legal analysis: on the resolution of conflict*, Berlin, Duncker und Humblot, 2000, 213p.

candidats est à même d'apporter une réponse à cette question. En outre, l'avenir de la profession d'avocat fait l'office d'un rapport qui formule également des propositions concernant la formation juridique et le rôle des universités⁸⁰. Cela nous conduit à évaluer le rôle des universités dans la contribution à la formation de bons magistrats.

Le grade académique joue également un rôle décisif sur le taux de réussite. Cette constatation semble à première vue attendue et logique : un grade académique facilite la réussite à l'examen. Cependant, l'examen a pour objectif d'évaluer la maturité du candidat et la présence de certaines compétences autres que les connaissances acquises lors de la formation académique. Faut-il donc abandonner cette supposition et conclure que l'examen d'aptitude professionnelle et le programme académique reposent sur les mêmes connaissances et compétences ?

Par contre, il se pourrait que l'obtention d'un grade académique et la réussite à l'examen soient déterminées par un même facteur, le facteur *g*. Celui-ci fait référence à l'intelligence générale d'un candidat. On sait que ce facteur est très prédictif de la performance professionnelle ainsi que du succès académique⁸¹. Il s'impose donc d'identifier si ce facteur *g* explique réellement le lien entre la réussite à l'examen et le grade académique. Suffirait-il alors d'évaluer l'intelligence générale des candidats juristes comme méthode de sélection ?

Finalement, nous sommes conscients du fait que les facteurs étudiés dans cette contribution ne sont peut-être pas tous appropriés pour l'étude de la validité conceptuelle de la méthode de sélection. D'autres facteurs, notamment le facteur *g*, le facteur général de personnalité⁸² de même que la conscience professionnelle⁸³, nous paraissent être tout aussi pertinents pour déterminer la validité conceptuelle de l'examen. La détermination de ces facteurs pourrait être nourrie par une conceptualisation du « juge idéal ». Quelles sont les qualités qui définissent ce juge idéal, et de quel façon sont-elles mesurables chez les candidats à une responsabilité de magistrat ?

80

https://justitie.belgium.be/sites/default/files/rapporttoekomstadvocatuurrapportavenirprofessionnel_avocatfinal.pdf

⁸¹ Y. GANZACH et P. PATEL, "Wages, mental abilities and assessments in larger scale international surveys: still not much more than *g*", *Intelligence* 2018, 1-7

⁸² D.H.M. PELT, D. VAN DER LINDEN, C.S. DUNKEL et M.P. BORN, "The general factor of personality and job performance: Revisiting previous meta-analyses", *International Journal of Selection and Assessment* 2017, 333-346

⁸³ M. OHME et H. ZACHER, "Job performance ratings: the relative importance of general mental ability, conscientiousness and career adaptability", *Journal of Vocational Behavior* 2015, 161-170