

論文式試験における段階評価に関する考え方（私案）

< 基本的な考え方 >

- ・ 新司法試験（以下、単に司法試験）の論文式は4科目であるが、実質的には、3つの法律基本科目については、各設問の中心的な問題がそれぞれ異なる法律分野に属する（たとえば、刑事系については、おそらく1問が刑事実体法中心、もう1問が刑事手続法中心となる）と考えられることから、段階式評価についても、合否の判定を含めて問題毎に行う。

この場合、3つの法律分野を含む民事系科目でも2問の出題が原則となっていることが若干問題となるが、この私案では、各問の法律分野が異なるという事情は同様であるという前提をとり、これも問題毎の段階式評価を行う。

選択科目については、未確定の部分が多いが、私案においては、問題数を1問と仮定する。

- ・ 以上の結果、4科目の試験であるが、実質的には、7問についてそれぞれ段階評価を行う。

科目ではなく問題毎に7科目の試験のように採点することは、新司法試験法の趣旨に反するのではないかと疑問も生じ得る。この点は、新司法試験法の趣旨の理解に依存するが、私自身は、科目別に試験を実施することと合否判定を問題毎に行うこととは別個の問題と考えている。得点評価方式の場合でも、結局のところ、各問の合計点で合否を決めることになると思われるし、短答式試験において、かりに現行試験と同様に科目毎の最低点を定めないとすれば、科目毎に試験を実施しても、科目毎の判定とはならず、試験実施の単位としての科目と合否判定の対象を同一に考えるべき必然性はないと考えられるからである。

また、上述のとおり、各問題は、本来の意味での融合問題を別とすれば、それぞれメインとなる法分野が異なるのが通常であると考えられ、実質的に見ても、科目毎の合否判定という考え方が、新司法試験法の趣旨と相容れないとはいえないであろう。

- ・ 段階をどの程度細分化するかは、考え方の分かれるところであるが、5段階ないし6段階程度が適当ではないかと考えている。ここでは、5段階評価を前提とする。Dの下をFとしたのは、不合格という意味を明らかにするためである。

- A とくに優秀な答案
- B 優秀な答案
- C 合格レベルを十分に超えている答案
- D 問題もあるが、一応合格レベルに到達している答案
- F 不合格答案

- ・ このランク分けをどの程度細かく行うかについては、さらに検討の余地があり、答案のバラつき方により、6段階、7段階程度の細分化もありうるかもしれない。しかし、得点評価方式が前提とする1点刻みの差別化は、資格試験としての性格を備えるべき新司法試験における評価方法として基本的な疑問があるだけでなく、各試験委

員の（場合によっては相当程度の）採点誤差を考慮すると、かえって不合理な結果を招くことにもなりかねない。

- ・ 段階評価においては、合格レベルに達しているかどうかの判断が最も重要であり、F判定は、単に5段階評価の最低ランクという以上の意味を持つ。総得点順に受験者を並べて一定の点数で合格ラインを決めるという発想ではなく、将来法曹となるにふさわしい一定のレベルをクリアしているかどうか、新司法試験の合否判断にとって最も重要な観点と考えられる。
- ・ 段階式評価に際しても、単純に試験委員の印象に委ねるのではなく、どのような項目にどのように配点をするか、基準を統一する。基準のレベルの統一については、各問題の試験委員の裁量に委ねるのではなく、試験委員全員が共通の理解をしておく。ただし、論文式試験では、事実の解析能力、法的分析力、構成力、文章の表現力、論旨の説得力など、個々の法的論点の評価に還元することが困難な評価要素が多く含まれることから、基準を統一する際にも、これらの点に十分に配慮することが必要である。
- ・ 段階評価方式は、あくまでランク分けを行うことに主眼があり、個別的な項目に関する評価は、最終的に段階評価を行うための評価「過程」にすぎない。

段階評価を行う際に、漠然と答案全体を読んで成績評価をするだけでは、印象のブレがきわめて大きくなるおそれが強いこと、とりわけ相当数の枚数を採点する際に、前後で採点基準に不一致が生じないようにする必要があること等を考慮し、段階評価を行う基礎となるデータとして、各試験委員は、統一された基準に照らして、まず点数化を行うことになる。

たとえば、採点に当たっては、100点満点をベースに考えて、ある評価項目に 点、ある評価項目に 点というような振り分けを行い、その合計点数を基礎として、 点以上と評価できる答案はAランクとするといったイメージを考えている。しかし、90点、88点、85点というような点数評価となった場合に、最終的には、それらの答案を同じランクとしてもよいかどうか（いいかえれば無視できる程度の相違と考えるか）、あるいは、両者は質的に異なったランクに位置づけられるべき答案であるかを判断することになる。90点と88点は無視できる差だが、88点と85点では、ランクが違うということになれば、たとえば前の2つはAランク、最後の1つはBランクとして評価が与えられる。各試験委員が採点結果として提出するのは、このランクのみであり、いわば段階評価に至るプロセスとしての素点は試験委員の内部参考資料にとどまる。

評価項目・基準の統一

評価基準をどのように定めるかを、新司法試験において出題される問題がどのようなものとなるかが明らかとされていない段階で具体的に例示することは容易ではないが、たとえば、長文のかなり複雑な事実関係を前提として、その法的解決の在り方を問うような問題をイメージすると、以下のような項目・基準を考え得る。

事実関係の分析能力

- ・ 客観的に確定している事実、当事者の単なる主張や真偽が明らかではない事実等を区別できているか。
- ・ 法的に意味のある事実とそうでない事実を区別できているか。

前提となる事実関係に即した法的分析が的確に行われているか

- ・ 考慮すべき法的問題点を適切に拾い上げているか。

- ・ 考慮すべき法的問題点について、従前の判例・学説の到達点や議論を正確に把握しているか。

実務的な観点に対する配慮

- ・ 訴訟における証明方法や証明の困難性等を考慮しているか。
- ・ 訴訟に関わる立場の相違が認識されているか（たとえば、訴訟代理人の立場から考えるべき場合と、裁判官として判断すべき場合、あるいは研究者が鑑定意見を述べる場合等の相違が考慮されているか）。

答案全体の論理的な構成力

- ・ 判例・学説を踏まえつつ、自分の考え方と結論を説得的に提示しているか。
- ・ 論旨全体が一貫したものとなっているか。前後の主張に不整合や矛盾がないか。
- ・ 結論の当否について、十分な考慮が払われているか。他の場合とのバランスなどが適切に考慮されているか。

表現力

- ・ 一読して理解しやすい、明快かつ達意の文章となっているか。
- ・ 不必要に冗長であったり、さらに詳しく説明すべき部分を省略した飛躍のある文章になっていないか。

これらの基準・項目は截然と区別できないものが少なくなく、また、各項目にどの程度ウエイトを置くかは、問題によって異なると考えられる。事実関係が複雑であれば、その基準によりウエイトが置かれ、法的分析がより中心となる問題であれば、そのウエイトがたとえば7割を超えるような場合もありうるかもしれない。いずれにしても、これらの項目についての評価を、厳密に1点刻みで行うことは新司法試験の趣旨からしても意味のある作業とはいえないと考えている。

< 具体的な合否決定の方式 >

論文式の評価

- ・ 2問についてF評価があれば、他の成績如何に関わらず不合格とする。この場合、短答式との総合評価は行わない。
- ・ Aは40点、Bは30点、Cは20点、Dは10点、Fは0点として点数化する。
民事系科目の評価については、中間報告案の配点比率に対応させて、それぞれ1.5を乗ずる。ただし、民事系の2問のうち、たとえば1問がもう1問の2倍の配点とされた場合には、前者の評価を2倍とし、後者はそのままの点数とすることが考えられる。

2名で採点することから、各問の最大点数は80点（民事系では120点ないし160点）となる。

得点評価方式とは異なり、満点を40点として、その点数の内訳を項目毎に考えるという発想ではなく、段階評価が優先し、Aランクと評価された答案に対して40点、Bランクに30点という考え方をとることになる。

得点化することの意味

段階式評価を得点に換算する必要性は、各科目によって配点が変わるという中間報告の立場に配慮する必要があること、短答式試験との総合評価を行う必要があること（後者は必然的に

点数化される)によるものである。40点, 30点という点数は便宜的なものであり, 4点, 3点という点数もありうる。

- ・ C評価が各問について十分に合格ラインに達しているレベルであることを考慮し, 論文式の合格ラインは, 合計点で160点(民事系を除く5問×20点+民事系2問×30点)×2委員=320点以上とする。

なお, 選択科目については, 合否判断のみを行い, 合格者は一律に20点(×2)を与えるという方法もありうるかもしれない。

選択科目間の難易度調整

各選択科目の受験グループのレベルがおおむね等質的であるという前提に立ち, かつ, 成績分布の割合を大まかな形でも目安として示すことになれば, 問題の難易度如何に関わらず, 成績分布のばらつきは, 大きく異ならないといえるかもしれない。したがって, 後述する成績評価の分布割合をどうするかがここでも重要なポイントとなる。

- ・ A評価をつける割合は, 同一採点グループの中で, 各試験委員につき1割程度(あるいは最大1割まで)とする。B評価は同じく2割5分程度(あるいは最大2割5分まで;これはA評価を含む割合。他も同じ)とする。もっとも, この割合は, 受験生の到達度に依存することから, これは一つの目安であり, 絶対的な基準ではない。絶対評価の考え方からいえば, このような割合を固定することにはむしろ問題もあるが, きわめて優秀ないし優秀な答案の割合はせいぜいこの程度の割合ではないかという推定に基づく例示である。

最大の問題は, 合格レベルを超えるとするC評価の割合をどのように設定するかである。段階評価方式の理念からいえば, 本来, この割合を設定することは矛盾であり, かりに受験者の8割が合格レベルに達していれば8割がC評価を受けるべきであり, 逆に, 2割しか合格レベルに達していなければ(たとえ司法修習の受け入れ体制に余裕があったとしても)それ以上の割合に対してC評価を与えるべきではないと考えられる。数年間新司法試験が実施されれば, 自ずからその割合も安定することになると予測される。

しかし, 当初は, この予想がつきにくいこと, グループ間における大きな不均衡を回避する必要があること(例えば, 同じ採点グループで一方の試験委員が3割をC以上とし, 他方の試験委員が6割をC以上とすることには問題がある)等を考慮すると, 一定の割合(たとえば6割~7割程度を限度とする)を設定する必要があるかもしれない。同じ問題は, D, Fについても存在する。これらは, 上述の評価基準をどの程度統一し, 試験委員がどの程度共通認識を有することができるかに依存するとともに, 何よりも, 各法科大学院が, 法科大学院における教育理念の根幹ともいえる厳格な成績評価と修了認定をどの程度実践するかにかかっている。

- ・ 試験委員の間で, 段階評価のランクづけが分かれた場合に, 1ランクの相違については, 原則として単純に点数を合算する。ただし, 1名がFをつけ, 他がD以上をつけた場合には, 他の試験委員1名がさらに評価し, DかFに統一する。

2 ランク以上評価が分かれた場合（たとえば，A と C ）には，同じく他の試験委員 1 名が評価し，評価を統一する。

これらの場合，最後の試験委員の意見が最終判断となるが，2 名の採点委員の評価結果が異なることを踏まえての判断であり，これはやむを得ないと考える（上の例では最後の試験委員の評価が D となる場合を想定していないが，これは実際にもあり得ないのではないか）。

1 ランクの相違について調整を行わないとした実際的理由は，このようなケースが頻繁に生じうるのではないかという懸念と，1 ランクの誤差は，平均で C ランクの成績を得ていれば合格判定を得ることができるとすれば許容範囲ではないかと考えた点にあるが，1 ランクでも違えば調整を行うべきだとの考え方もありうる。とくに C と D 判定が分かれるような場合には，調整の必要性が高いかもしれない。

評価が D と F に分かれた場合，いずれかに統一するという趣旨は，F 評価が単なる総合点の問題でなく，2 問について F があれば，他の成績を問わず全体として不合格となるという考え方と連動している。

また，2 ランク以上相違する評価については，第 3 の試験委員が評価を統一するとしていた点については，たとえば A ， C と分かれていた場合に，この「幅」の中で評価を統一するという趣旨である。したがって，今の例では A ， B ， C のいずれかに統一されることになる。第 3 の試験委員は，ランク付けについて評価が分かれていたことを前提とした上で，それをチェックするという見地から採点を行うのであり，まったく白紙で採点をするのとは異なるという考え方が前提となっている。

短答式試験との総合評価

- ・ 段階式評価を以上のように考える場合，短答式試験との総合判断の必要性は乏しいと考えるが，新司法試験法の枠組みからいえば，単純に，短答式・論文式それぞれの合格ラインを充たしていれば，それで最終合格とすることに問題があるかもしれない。
- ・ 一案として，論文式合格点数に若干足りない者，たとえば総点が 280 点以上 320 点未満の者については，短答式合格ラインを超える点数を加算することにより，合計点で論文式試験と短答式試験の合格点数の総点を上回る場合には，総合評価で合格とする。

単なる救済方法に過ぎず，あまり合理的な制度とはいいいくいが，このような方法をとる場合には，論文式の最低ラインは 280 点であるが，合計では 320 点 + 短答式合格点数に達していなければならないという言い方をすべきかもしれない。

得点と段階を併用すること（「合格ラインを 320 点以上」とし，かつ，「F 評価が 2 つ以上あれば不合格」とすること）により，総得点でのより高得点者が不合格となり，より低得点者が合格となる現象が生じうるが，これは，段階評価方式をとる場合に，不合理な結果であるとはいえないと考えている。

ここでも重要であるのは，新司法試験における合否判定の目的は，司法修習を経て，法曹となるにふさわしい能力を備えているかどうかにある。各問題について，いわば合格ぎりぎりのレベルであっても，それをすべて超えていれば合格させるべきである。他方，たとえば，民事系に関する成績は優秀であっても，刑事系についてはまったく不十分な理解しか得ていない場

合には、全体としてみれば、法曹としての適格を充たしていないと考えられる。

この点は、法科大学院の修了判定についてもまったく同様のことがいえる。必修科目とされる法律基本科目等については、それらのすべてについて、合格判定を得てはじめて法科大学院を修了することができる。多数の科目できわめて優秀な成績を収めていても、特定の必修科目について合格しなければ、修了要件を充たすことができず、したがってまた新司法試験を受験する資格を得ることもできない。

論理的には、1問でもFがあれば他の成績を問わず不合格とするという考え方も成り立ち、その方が一貫しているといえるかもしれないが、1問の失敗が決定的となること、法科大学院の修了判定（ないし予備試験合格判定）を経ていることから、1問の失敗については、他の成績次第でなお許容範囲であるとするのが私案の立場である。

評価の不一致の場合の調整と試験委員の負担

採点結果について、場合によって、さらに第三者による再評価が行われることになると、試験委員の負担がさらに大きくなるという懸念も考えられる。しかし、試験委員は、修習資格を与えるべきかどうかというきわめて重要な判定に関与する以上、それなりの負担と責任を引き受けることはやむを得ないものと考えられる。試験委員の過剰な負担を避けるためには、各試験委員の採点枚数をどの程度とするか、必要な試験委員の数と質をいかに確保するかを検討することが不可欠である。