【原著】 大学入試研究ジャーナル第 32 号，92-97，2022

ペーパー・インタビューの評価基準の作り方について

石井 志昂，吉村 宰（長崎大学）

ペーパー・インタビュー（面接に代わる筆記試験）の評価基準のよりよい作成方法を探るため，高校生を対象とした試行試験で得られた答案を，2 通りの評価基準で採点し結果の一貫性を検討した。評価基準はいずれも 5 段階で，1，3，5 の各段階に該当する答案の特徴が書かれている。大学院生 5 名に採点を依頼し，終了後，評価基準に関する意見を聴取した。得られた意見を参考に評価基準を修正し，同様に別の大学院生 5 名に採点の依頼，及び意見を聴取した。その結果，各評価段階（1～5）に該当する答案の特徴が具体的すぎると採点が困難となること，採点前に採点者同士で評価合わせを行うことで，得点のばらつきが抑えられることが明らかとなった。

キーワード：主体性等評価，一般選抜，構造化面接，ペーパー・インタビュー，評価基準

1 問題と目的

入学者選抜での「学力の 3 要素」を多面的・総合的に評価することの具体的な方策として，高大接続システム改革会議の「最終報告」（2016 年 3 月 31 日）では，調査書の利用，面接，プレゼンテーション，集団討論の実施などが挙げられているが，募集人員が大きな一般選抜では実施面での実現性に乏しい。

その中，長崎大学では，すべての入試区分において調査書の活用，及び面接によって「主体性を持って多様な人々と協働して学ぶ態度（主体性等）」の評価を行うことを原則とした上で，志願者数の規模から 1 日での実施が不可能な場合に，面接に代わる「ペーパー・インタビュー」を行うこととした。ペーパー・インタビューとは，構造化面接試験を紙面上で行おうとするものであり，事前に設定した「評価する力や特性」に関する問題文に対して，自らの考えや経験等を試験時間中に文章で解答する筆記試験である。吉村・石井（2021）は，高校生 168 名に対しペーパー・インタビューの試行試験を実施し，答案に対する高校教師と大学教員の評価について検討を行った。その結果，高校教師と大学教員の評価はある程度一致していたが，採点者によるばらつきが大きいことを報告している。

本研究は，ペーパー・インタビューの評価基準に着目し，よりよい評価基準の作り方を探索的に検討するものである。

2 試行試験の詳細

吉村・石井（2021）の試行試験には，県内 5 校の高校 2 年生計 168 名が参加した。ペーパー・インタビューの問題は 2 問あり，各問題における評価する力や特性は，問題 1 が「困難を突破する力（定義：やるべき事がうまくいかない時に，あきらめず色々な方法を試し，自力で最後までやり遂げようとする，課題解決への執着心）」，問題 2 が「企画・運営力（定義：物事を企画しそれを運営する力）」である。表 1 は試行試験にて出題した問題である。回答時間は各問題 45 分とした。答案用紙は B4 判片面に横書き 30 行の解答欄を印刷したものであり，文字数に制限を設けなかった。

答案の採点は評価基準に従って行った。採点者間における採点結果のばらつきを抑えるため，各評価段階に該当する答案の特徴を具体的に示した評価基準を事前に作成した。図 1 に，1 回目の採点に用いた問題 1 の評価基準を示す。具体的には，図 1 の評価 5 に対応する大枠の評価基準が「諦めずいろいろな方法を試し何がなんでも最後までやり遂げようとする。」であり，枠内にある各項目が，評価 5 となる答案の特徴を示している。問題 2 についても同様の評価基準を作成した。図 1 の評価基準は得点上 5 段階評価だが，実質的には 3 段階となっている。5 段階評価にすると，得点の集中する評価 3 の中で，「3 の上」「3 の下」など評価が細分化し，評価 4 や評価 2 と差別化できなくなってしまうのを防ぐためである。

3 1 回目の採点

1 回目の採点では，試行試験で収集された高校生 168 名の答案をアルバイトとして募集した大学院生 5 名が，図 1 の評価基準に従い採点を行った。ただし，問題 1 は 168 名の内 1 名が英語のみで回答を行っていたため，167 名分を採点対象とした。答案の順序効果を防ぐため，ランダムに並べ替えた答案を 5 セット用意し，採点者は独立にそれぞれに割り当てられた答案を採点した。採点に際して，評価基準に従って採点を行うこと，採点に関する質問は受け付けないこと，

表 1 試行試験で出題したペーパー・インタビューの問題文

【問題 1】
高校入学以降，やらなければならないのに難しくてなかなかうまくいかず困った，というような経験を思い出してください。それはどのようなことでしたか。どんな些細なことでも構いませんので，以下の点を踏まえながら，できるだけ詳しく具体的に説明してください。複数思いつく人はその中からどれか 1 つを選んでください。 そのような経験を思いつかない人は，なぜ自分にはそのような経験がないのかについて書いてください。 ○それはいつ頃のどのような出来事だったか。 ○なぜそれをやらなければならなくなったのか。また，そのことについてどう思ったか。 ○どのような点が難しかったか。 ○どのように対応すればよいと思ったか。それは簡単に実行できたか。 ○簡単には解決できず困った時にどう思ったか。またどのようにして困難を解決しようとしたか。 ○最終的にうまくいったか。うまくいった場合なぜうまくいったのか。うまくいかなかった場合なぜうまくいかなかったのか。 ○もし次に同じような場面に直面したらどのように行動するか。その理由はなぜか。
【問題 2】
高校入学以降，文化祭，体育祭，学級行事，部活，学外の活動，友人同士の集まりなど何らかの活動の計画，または運営，あるいはその両方で何か頑張ったことを思い出してください。それはどのようなことでしたか。どんな些細なことでも構いませんので，以下の点を踏まえながら，できるだけ詳しく具体的に説明してください。複数思いつく人はその中からどれか 1 つを選んでください。 そのような経験を思いつかない人は，なぜ自分にはそのような経験がないのかについて書いてください。 ○いつ頃のどのような内容の出来事だったか。 ○どういういきさつでその活動に関わることになったのか。 ○あなたは具体的に何をしたのか。 ○計画や運営の際にどのような工夫を行ったか。 ○なにか困ったことが起こったか。起こった場合，その原因はなんだと思うか。それにどのように対応したか。起こらなかった場合，どのような問題が起こっていたら対応に苦しんだと思うか。 ○最終的にうまくいったか。なぜうまくいったのか，あるいはなぜうまくいかなかったのか。 ○結果について満足したか。どのような点で満足いったか。どのような点に満足がいかなかったか。 ○反省点はあるか。もし次に同じような経験ができるとしたら何に気をつけたいと思うか。その理由はなぜか。

採点者間で答案や採点の話をしないことを注意事項として伝えた。採点終了後，採点過程そのもの，及び評価基準について約 1 時間半のインタビューを行った。

表 2 に採点者 5 名（A～E）による採点結果の度数分布を示す。表から各問題とも採点者によってばらつきが大きいことがわかる。

問題 1 では，採点者 D が 67 名に対し評価 1 をつけ，他の採点者よりも厳しい評価を行っている。対して，採点者 C は 94 名に評価 4 と 5 をつけており，採点者による評価の甘い辛いは大きく異なっている。

表 3 には，5 名の採点者のうち，2 名ごとの得点一致割合を示す。一致割合算出にあたり，採点者間の採点結果の差の絶対値が 1 以下のケースを「一致」，絶対値が 1 を超えるものを不一致とした。不一致は，例えば採点者（A, B）のペアを例に取ると，A-B >1 であれば「上」，A-B < −1 であれば「下」のように 2 つのケースに分けた。さらに全答案に占める「一致」の割合を「一致割合」とし，表にまとめた。

表 3 から，問題 1 は問題 2 と比較すると，（C-D）で 38.9%，（A-B）で 94.3%などと一致割合のばらつきの大きさが目立つ。一方，問題 2 では極端に低い一致割合は見られなかったが，最も高いもので 82.7%と，全体的に一致割合は問題 1 に比べ低いと言える。5 名の採点者間信頼性を検討するため，級内相関係数（Intraclass Correlation Coefficient：ICC）の算出を行った結果，問題 1 は ICC（3, 1） = 0.460（95%CI：0.399 – 0.523），問題 2 は ICC（3, 1） = 0.433（95%CI：0.373 –0.497）であった。

困難を突破する力

やるべき事がうまくいかない時に，諦めず色々な方法を試し，自力で最後までやり遂げようとする，課題解決への執着心。

当てはまると判断すれば，躊躇なく評定を5や1としてください。

5 (優)	諦めずいろいろな方法を試し何がなんでも最後までやり遂げようとする。 • 自力で課題を解決しようとしている。 • 課題の把握と分析ができている。 • うまくいかなくても諦めず，なんとかして課題を解決しようとしている。 • 課題解決の方法を自分で考えており，独自性がみられる。 • いろいろなソースを調べたり，調査したりして，課題解決方法を探っている。 • 入学後も，何事もあきらめず最後までやり遂げることが期待できる。
4 (3+)	3 を上回るが 5 には届かない。
3 (並)	まじめに課題に取り組むが，困難な場面では安易な方法で乗り切ろうとする。 • 課題を解決する意欲はあり，課題を解決する方法を考えている。 • 課題解決の工夫が，先生，友人を頼る，ネットから引用するなど比較的安直である。 • 課題を解決する意欲はあるが，課題の解決に執着する様子は見られない。 • 課題が，学校から与えられるような皆がやらなければならないものである。
2 (3-)	1 を上回るが 3 には届かない。
1 (劣)	具体的な経験をあげられない。もしくは具体的な経験をあげるが， • なぜ難しかったのかがわかっていない。 • うまくいかなかったことの理由に自分の努力や工夫の不足をあげてない。 • なぜうまくいったか（うまくいかなかったか）の理由がわかっていない。 • 困難を避けることで解決している。他人任せである。 • 簡単に諦めている。

図 1 採点 1 回目の評価基準表（問題 1）

採点後のインタビューは，採点者に（1）評価してみてどう思ったか，(2) 評価基準に問題があると思うか，(3) 評価基準をどう改善すればよいと思うか，(4) その他気づいたこと，について尋ねた。

「(2) 評価基準に問題があると思うか」について，図 1 の評価基準で定めた各評価段階の具体的な項目が，チェックシートのようになってしまい，評価 1 の項目と評価 3 の項目両方に該当するような答案があった場合にどの得点を与えるのが難しかったという意見が得られた。具体的な項目を箇条書き的に定めると，1 個でも当てはまれば評価 5 なのか，何個当てはまればいいのかなど評価の判断が難しくなってしまうことがわかる。

表 2 採点 1 回目の採点結果の度数分布表

採点者	問題 1「困難を突破する力」					問題 2「企画・運営力」				
	1	2	3	4	5	1	2	3	4	5
A	11	36	79	26	15	23	19	72	22	32
B	10	36	65	49	7	40	49	61	16	2
C	11	13	49	44	50	22	13	73	35	25
D	67	52	43	4	1	45	47	55	14	7
E	10	42	68	39	8	9	33	70	37	19

「(4) その他気づいたこと」は，答案に対して「もう少し書けていて欲しい」などの意見がみられた。採点した答案は，吉村・石井（2021）で述べた通り高校 2 年生が対象で参加のインセンティブもないものであった。経験したエピソードに関しては記述できているが，それに対する振り返りや，経験から得られたものなどに関する記述が不十分な答案が多かったとの指摘もあった。

1 回目の採点結果から，評価基準の各評価段階に該当する答案の特徴を具体的に定め箇条書きで表現すると，採点時にチェックシートのように使われてしまい，採点に困難が生じることが明らかとなった。

4 評価基準の修正

評価基準修正の方針として，どの段階にあてはまるかの判断が難しくなるのを防ぐため，各評価段階に該当する答案の特徴を具体的に定めるのではなく，答案固有の記述によらない大枠での評価基準にすることと

表3 採点1回目における採点者2名ごとの採点結果一致割合

組み合わせ	問題1「困難を突破する力」				問題2「企画・運営力」			
	下	一致	上	一致割合	下	一致	上	一致割合
A-B	6	156	5	93.4	5	124	39	73.8
A-C	34	127	6	76.0	19	130	19	77.4
A-D	1	103	63	61.7	6	112	50	66.7
A-E	9	149	9	89.2	19	134	15	79.8
B-C	26	137	4	82.0	39	128	1	76.2
B-D	1	109	57	65.3	17	139	12	82.7
B-E	6	155	6	92.8	37	130	1	77.4
C-D	1	65	101	38.9	4	113	51	67.3
C-E	1	134	32	80.2	23	129	16	76.8
D-E	56	109	2	65.3	48	116	4	69.0

した。図2に修正後の評価基準を示す。

また，2 回目の採点に先立ち，複数枚の答案を採点者間同士で採点し，採点結果の甘い辛いや評価基準の理解について，評価合わせを行った。

表 2 のペーパー・インタビューの問題用紙には，解答の際にふまえるべき点が記されている。修正後の評価基準では，各問題に示されているふまえるべき点をそれぞれ以下の 4 つに分けた。問題 1 は「困難を感じたエピソード」「解決の方法」「取り組みの評価，その結果となった理由の考察」「今後について」，問題 2 は「企画・運営をしたイベントのエピソード」「イベントの結果と，その結果になった理由」「反省点とその理由」「今後について」とした。

修正後の評価基準において評価 3 以上は，各問題とも 4 つの内容がすべて記述されていることを前提にして評価を行う。評価 3 と評価 5 との評価のポイントは，自らの経験や取り組みを客観視できているかどうかである。

評価 3 は，問題文の指示に従い，経験したエピソードや取り組みの評価などの記述がなされているが，平均的な内容の答案を想定している。評価 5 は，エピソードを客観的にとらえることができ，相手を意識

困難を突破する力

やるべき事がうまくいかない時に，諦めず色々な方法を試し，自力で最後までやり遂げようとする，課題解決への執着心。

当てはまると判断すれば，躊躇なく評定を5や1としてください。

評価	評価基準
5	課題解決過程が俯瞰できており，自らの力で困難を突破しようとし続ける態度が明確かつ容易に読み取れる記述である。今後も直面する困難に対し，同様の態度で臨むことが期待できる。
4	エピソードが平凡すぎるものは4とする（学校の課題提出，早寝早起き，徹夜など）
3	経験した困難とそれを解決しようとしたエピソードを具体的に記述している。解決方法が容易に思いつく（単語帳を作る，時間の有効活用など）ものや，取り組んだ結果の評価・理由の考察が浅い。
2	解決方法や工夫が「もっとがんばる」「たくさんする」であるもの，取り組んだ結果の評価・理由の考察が欠けているものは2とする。
1	記述されたエピソードが，困難を感じた出来事のみの記述にとどまっているため，困難解決の過程が全く評価できないもの。
0	経験がないと回答。困難から回避しようとしている。記述すべき内容を理解していない。困難を突破しようとした経験を記述していない。箇条書きなどひとまとまりの文章として成立していない。

※答案は問題文の指示に従って，「困難を感じたエピソード」，「解決の方法」，「取り組みの評価，その結果となった理由の考察」，「今後について」の4つの内容に分かれているものを前提としています。評価5と評価3は，どちらも4つの内容がすべてそろっているうえで，内容について客観視できている（自分の目線だけではなくメタ的な視点を有している）かどうかを判断のポイントとしてください。評価3と評価1は，上記4つの内容のうち「困難を感じたエピソード」以外でどの記述がされているか判断のポイントとしてください。評価4は評価5に達しない答案，評価2は評価3に達しない答案であると判断してください。

図2 修正後の評価基準（問題1）

表 4 採点 2 回目における採点結果の度数分布表

採点者	問題 1「困難を突破する力」						問題 2「企画・運営力」					
	0	1	2	3	4	5	0	1	2	3	4	5
a	6	9	46	71	28	7	5	2	24	81	48	8
b	2	9	42	86	17	11	4	2	25	74	41	22
c	18	5	11	95	8	30	15	1	27	69	37	19
d	8	17	48	60	27	7	11	27	43	57	17	13
e	8	14	44	91	7	3	15	12	35	56	39	11

した明確かつ容易な記述である答案とした。

評価 3 と評価 1 は，エピソード以外の内容についてどれが記述されているかを評価のポイントとし，評価 1 はエピソードのみ記述されている答案とした。

また，図 1 で示した 1 回目の評価基準では，問題で問われている内容に対して「経験がない」と解答した答案についての評価について基準を設けていなかった。修正後の評価基準では，評価 0 として，「経験がない」「ひとまとまりの文章として成立していない」などの答案に対する評価段階を追加した。

5.2 回目の採点

修正後の評価基準について評価するため，1 回目の採点とは別の大学院生 5 名に同じ手続きで答案の採点を依頼し，その後インタビューを行った。5 名の内訳は，同一の医歯薬系専攻に所属する修士課程 1 年生 4 名，及び博士課程 1 年生 1 名で，採点に関わる経験を有する者はいなかった。ただし 2 回目の採点は，1 回目とは異なり事前に評価合わせとして問題 1 と問題 2 を各問題 5 枚ずつ話し合いながら採点する時間を設けた。

表 4 に採点 2 回目の採点者 5 名（a～e）における採点結果の度数分布表を示す。問題 1 は大半が評価 2～3 に該当しているのに対し，問題 2 は評価 3～4 の間に該当している。採点者 c における問題 1 の採点結果は他の採点者と異なり，0 や 5 と採点した答案が多いことがわかる。特に問題 1 では，採点者 b と 0 のつけ方に大きな違いがみられた。

表 5 には採点 2 回目における採点者 2 名ごとの得点一致割合を示す。また，図 3 に採点 1 回目と 2 回目における一致割合を問題ごとに点で示した。ただし，マーカーが重なるのを防ぐため幅を散らして描画している。図 3 より，問題 1 に関しては，評価合わせを行い，修正後の評価基準を用いて採点を行った採点 2 回目の一致割合のほうが，ばらつきが抑えられていることがわかる。問題 2 は 1 回目の採点よりもばらつきが広がっているが，一致率が 8 割を超えている組み合わせは増えている。採点者 2 名ごとの採点結果の差を確認すると，問題 1 では最大で 4（該当答案数：3），問題 2 は最大 5（該当答案数：2）であった。採点者間信頼性は，問題 1 で ICC（3, 1） = 0.509（95%CI：0.449 – 0.569），問題 2 で ICC（3, 1） = 0.451（95%CI：0.390 –0.514）であった。

採点後のインタビューも 1 回目同様に行った。2 回

表 5 採点 2 回目における採点者 2 名ごとの採点結果一致割合

組み合わせ	問題 1「困難を突破する力」				問題 2「企画・運営力」			
	下	一致	上	一致割合	下	一致	上	一致割合
a–b	13	143	11	85.6	10	146	12	86.9
a–c	22	131	14	78.4	13	139	16	82.7
a–d	10	142	15	85.0	8	115	45	68.5
a–e	3	148	16	88.6	9	126	33	75.0
b–c	25	122	20	73.1	9	137	22	81.5
b–d	8	143	16	85.6	9	109	50	64.9
b–e	1	151	15	90.4	11	119	38	70.8
c–d	11	129	27	77.2	10	126	32	75.0
c–e	7	127	33	76.0	9	136	23	81.0
d–e	4	152	11	91.0	22	136	10	81.0

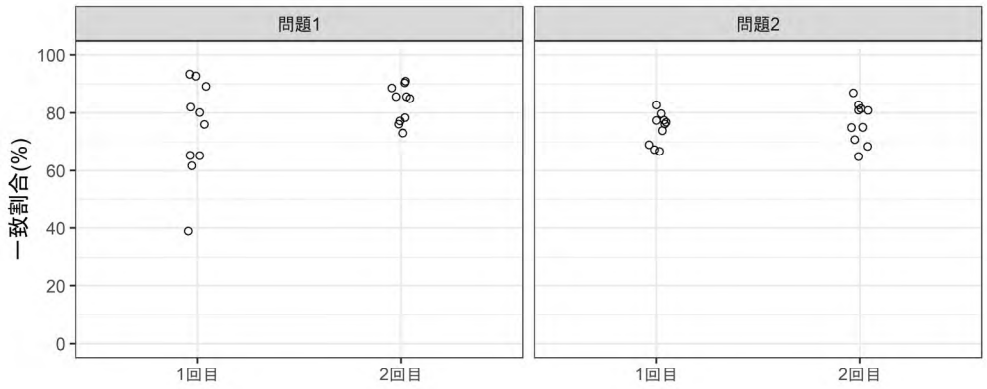

図3 採点1回目と2回目の一致割合

目では事前の評価合わせについての意見も聴取した。

どのように採点したかについては「一応反省は書かれてあるが，一文だけで反省というより感想なんじゃないかと思い，評価 2 と 1 で悩んだ」「(問題 2「企画・運営力」について）高校生の経験としては難しい内容だったのかなと思った，経験がないと書いた答案や，経験がないから全く違うことを書いているような答案が結構多かった」などの意見が得られた。

事前の評価合わせについては，「(問題 1 は部活動の内容が多く）部活動をしていない自分からしたら新鮮なエピソードに感じ，高く評価をつけることがあったが，周りの人に聞くと結構普通の話であることもあり，新鮮に感じないように調整した」「(「平凡」という評価基準表について）個人の主観によるものなのでそこでずれが生じたのではないか」など，評価合わせの時の採点結果のずれについて意見が得られた。ただし，評価合わせでは，個々の答案に何点与えるかという議論は行われたが，評価基準の解釈や意思統一についての議論は行われなかったということである。

6 まとめ

本研究は，面接に代わる筆記試験であるペーパー・インタビューの評価基準をどのように作成すればよいかについて探索的に検討し，以下の知見が得られた。

(1) 具体的な特徴を箇条書きにする評価基準では，特徴がチェックボックスのように用いられ，評価が困難となることがある，(2) 評価基準の記述を包括的なものとし，順序がつけられるような判断基準を明示するとそのような困難は見られなくなる，(3) 評価合わせでは特定の答案に与える点数を議論するのではなく，評価基準の解釈について議論することが重要である，(4) 評価基準の形式によって採点者による採点結果のばらつきが小さくなる可能性がある。

ペーパー・インタビューに限らず，主観を伴う評価では採点者の経験や価値観，評価基準の解釈が一致しない。例えば本研究で言えば，改善後の評価基準における問題 1 の評価 4「エピソードが平凡である」の「平凡」の解釈が採点者間で異なった。解決策としては，解釈の余地ができるだけ小さくなるように留意しながら評価基準を作成する，どのように解釈するかを事前に採点者間で決めておく，などが考えられる。

加えて，問題 2「企画・運営力」については問題 1 よりも経験がないと解答した答案が多く見られたとの意見があった。つまり用意した評価基準が答案のレベルと一致しなかったということである。評価基準を作成する際にはどのような答案が得られるかを想定しておく必要がある。ある種の「難易度」を考慮した問題作成を行うことが重要であるとも言い換えられる。

本研究ではペーパー・インタビューのよりよい評価基準の作り方をテーマとしたが，考慮すべき共変する要因が複数あるため，一般的にこれが最善であるという解は得られず，採点者も大学院生であるため，実際の入試採点業務に結果をそのまま当てはめることは注意が必要である。しかし本研究の結果は，評価基準は作題後に作成するのではなく「問題，予想される答案」とともに一体的に作成する必要があることを示唆するものである。その具体を明らかにすることを今後の課題としたい。

参考文献

高大接続システム改革会議（2016年3月31日）.「高大接続システム改革会議『最終報告』」 文部科学省

吉村幸・石井志昂（2021）.「ペーパー・インタビューの試行結果について――面接に代わる筆記試験の有用性の検討――」『大学入試研究ジャーナル』**31,** 161－166.