

PENINGKATAN KUALITAS PENILAIAN INSYA'

Asyraf Muzaffar

Universitas Islam Negeri Ar-Raniry Banda Aceh

Abstrak

Penilaian insya' memiliki kelemahan dalam hal atribut dasar penilaian tes yang baik, yaitu reliabilitas dan validitas yang rendah. Kelemahan ini memang merupakan bagian inheren dari penilaian insya' seperti tes subjektif lainnya. Namun demikian, keadaan ini semakindiperburuk jika pemeriksa yang memeriksa insya' kurang memahami permasalahan penilaian insya' dan kurang menguasai metode-metode penilaiannya. Oleh karena itu perlu adanya kemauan dan upaya pada setiap pemeriksa untuk meningkatkan pemahaman terhadap masalah ini. Dalam pelaksanaan penilaian, seorang pemeriksa akan memutuskan untuk menggunakan metode yang paling sesuai dengan mempertimbangkan keadaan yang melingkunginya. Dalam konteks pembelajaran, al-thariqah al-jama'iyah dan al-thariqah al-tahliliyyah atau penggabungan keduanya akan memberikan manfaat yang lebih aplikatif bagi pencapaian tujuan pembelajaran. Di sisi lain al-thariqah al-inhiba'iyah atau penggabungannya dengan al-thariqah al-jama'iyah juga memiliki keunggulan tersendiri berdasarkan pandangan bahwa insya' yang baik lebih dari sekedar kumpulan unsur-unsur individual. Pemeriksa akan mampu menangkap kesan keseluruhan dan keutuhan sebuah insya' jika menggunakan al-thariqah al-inhiba'iyah. Karena itu pemeriksa harus menyadari dan mengetahui kelebihan masing-masing metode sehingga dapat membuat pilihan yang paling sesuai dengan level pendidikan dan tujuan pembelajaran yang telah dirumuskan.

Kata kunci: Penilaian insya', al-thariqah al-inhiba'iyah, al-thariqah al-jama'iyah dan al-thariqah al-tahliliyyah, al-thariqah al-aaliyah

A. Pendahuluan

Tes *insya'* (Inggris: *composition*), sebuah kategori tes subjektif yang lazim digunakan untuk mengukur *maharah kitabah* dalam mata pelajaran Bahasa Arab, menghadapkan guru atau dosen yang memeriksa jawaban murid atau mahasiswa terhadap tes tersebut kepada tantangan reliabilitas, validitas, konsistensi dan objektifitas

yang sangat sulit. Adalah penting bagi seorang pemeriksa untuk menyadari tantangan ini serta memahami, menguasai pendekatan-pendekatan dan teknik-teknik yang disarankan dalam buku-buku evaluasi pembelajaran atau teori tes dan memodifikasinya sesuai dengan tujuan dan penekanan yang dia pilih agar dapat meningkatkan kualitas hasil pemeriksaan jawaban. Meskipun kecilkemungkinannya untuk bisa mencapai reliabilitas, validitas, konsistensi dan objektivitas yang tinggi pada hasil pemeriksaan *insya'*, ada beberapa upaya yang dapat dilakukan guru atau dosen untuk mereduksi subjektivitas dan meminimalisir inkonsistensi hasil pemeriksaan, dan sebaliknya meningkatkan reliabilitas, validitas, konsistensi dan objektivitas.

Tulisan ini dimaksudkan untuk mendiskusikan upaya-upaya tersebut dalam kerangka pendekatan-pendekatan dan teknik-teknik pemeriksaan jawaban *insya' (tahririy)* yang telah disarankan oleh para ahli dan praktisi pengukuran hasil pembelajaran. Tulisan ini mencoba mengelaborasi pendekatan-pendekatan dan teknik-teknik yang disarankan tersebut ke dalam bentuk yang lebih aplikatif sehingga dapat diterapkan secara lebih terstruktur dan terukur. Diharapkan para guru dan dosen dapat terbantu dalam memilih dan menerapkan pendekatan dan teknik pemeriksaan yang memungkinkan mereka meningkatkan kualitas hasil pemeriksaan jawaban *insya'* sehingga sebagian tantangan reliabilitas, validitas, konsistensi dan objektivitas dapat ditanggapi dengan lebih baik.

B. Pendekatan Pemeriksaan Insya'

Walaupun sebagian kemampuan (*sub-skill*) *insya'* dasar seperti *imla'* dan tanda baca dapat dites dengan tes objektif, kemampuan utama *insya'* yang lebih tinggi tidak dapat diukur dengan tes objektif. Oleh karena itu pengukurannya hanya bisa dilakukan dengan tes subjektif, yaitu dengan meminta siswa untuk membuat sebuah tulisan (*insya'*). Berbeda dengan tes objektif yang memiliki karakteristik *dichotomous* dimana kualitas jawaban murid atau mahasiswa terhadap sebuah butir soal dapat dipastikan masuk dalam hanya satu dari dua kemungkinan: benar atau salah, jawaban tes subjektif seperti *insya'* memungkinkan adanya jawaban-jawaban yang kualitasnya terentang dalam kontinum benar hingga salah. Hal ini berarti bahwa jawaban tes *insya'* memiliki variasi kualitas kebenaran jawaban yang jamak (*polytomous*). Seorang

pemeriksa akan menghadapi jawaban-jawaban insya' yang mungkin tidak sepenuhnya benar tetapi hampir mendekati sempurna, tidak sepenuhnya salah tetapi kualitasnya sangat rendah, kualitas kebenaran jawaban sekitar lima puluh persen atau proporsi kebenaran lainnya. Keadaan ini memberikan tantangan yang tidak mudah bagi seorang pemeriksa untuk menentukan kebenaran sebuah jawaban secara proporsional dan menentukan kualitas relatif jawaban-jawaban secara komparasional. Pada sisi lain, di samping masalah kualitas jawaban, ancaman terhadap kualitas hasil pemeriksaan ditambah lagi dengan preferensi subjektif pemeriksa yang semakin menjauhkan hasil pemeriksaan jawaban insya' dari kualitas pemeriksaan yang baik.

Muhammad Ali al-Khau¹ dalam bukunya *al-Ikhtibar al-Lughawiyah* menyebutkan ada empat metode yang dapat diikuti dan diterapkan dalam memeriksa insya'. Yang pertama adalah *al-thariqah al-inthiba'iyah* atau Metode Kesan. Dengan metode ini, pemeriksa membaca insya' murid dan menentukan skornya berdasarkan kesan umum yang dia dapatkan dari insya' tersebut. Metode ini memungkinkan pemeriksaan dilakukan dengan cepat, tetapi tingkat reliabilitas yang dapat dicapai lebih rendah dibandingkan dengan metode-metode lainnya.

Yang kedua adalah *al-thariqah al-jama'iyah* yang dapat diterapkan dengan melibatkan lebih dari satu orang pemeriksa yang secara mandiri memberikan skor masing-masing untuk setiap insya' murid. Dalam metode ini, skor seorang murid merupakan skor rata-rata yang dihitung dari skor-skor yang diberikan oleh para pemeriksa.

Yang ketiga adalah *al-thariqah al-tahliliyyah* atau metode analisa. Pemeriksaan jawaban insya' jika menggunakan metode ini dilaksanakan dengan terlebih dahulu menyiapkan atau menentukan unsur-unsur utama yang akan menjadi fokus penilaian insya'. Unsur-unsur yang dinilai tersebut, misalnya, terbagi kepada tata-bahasa, kosa kata, gaya bahasa, tanda baca, organisasi karangan dan lain-lain. Kemudian untuk setiap unsur ditentukan rentang nilainya, misalnya ١-٥ atau ١-١٠. Agar lebih mudah mempedomani

¹ Al-Khau¹, Muhammad Ali, *al-Ikhtibar al-Lughawiyah*, (al-Urdun: Dar al-Falah li al-Nasyr wa al-Tauzi', ٢٠٠٠), hal. ١٦٧-١٦٩

dalam proses pemeriksaan, unsur-unsur tersebut dapat dituangkan dalam sebuah tabel yang menyediakan kotak-kotak (*cell*) yang akan diberi tanda centang oleh pemeriksa sebagai *sub-score* unsur terkait.

Unsur	١	٢	٣	٤	٥	Nilai/u
Mufradat						
Qawaid						
Imla'						
Tanda baca						
Uslub/kelan						
Relevansi						
Total						

Gambar. ١. Tabel pedoman pemeriksaan Insyā' dengan *al-thariqah al-tahliliyyah*

Yang keempat adalah *al-thariqah al-aaliyah*. Dalam metode pemeriksaan insyā' ini pemeriksa membaca jawaban insyā' dan menandai kesalahan-kesalahan yang terdapat dalam jawaban tersebut. Pemeriksa kemudian menghitung jumlah kesalahan yang terdapat dalam jawaban dan berdasarkan skor yang ditetapkan untuk setiap kesalahan, maka total skor kesalahan dapat ditentukan. Pemeriksa selanjutnya mengurangi sejumlah total skor kesalahan tersebut dari skor maksimum jawaban insyā' yang dapat dicapai seorang murid jika jawabannya sempurna tanpa salah. Dengan mendasarkan pada pemberian skor melalui pengurangan skor jawaban salah dari skor maksimum, *al-thariqah al-aaliyah* ini memungkinkan tercapainya reliabilitas yang baik tetapi cenderung memberikan validitas yang rendah.

Dalam teori tes, validitas merupakan konsep yang sangat penting yang mencakup penilaian menyeluruh terhadap kualitas sebuah tes yang menjadi dasar pengambilan kesimpulan dan

tidak. Samuel Messick, seperti dikutip dalam Linn, mendefinisikan validitas sebagai "... an overall evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions on the basis of test scores or other modes of assessment."¹ Atau dalam ungkapan yang lebih sederhana, validitas sebuah tes merujuk kepada penilaian empiris dan teoritis sejauh mana kesahihan penarikan kesimpulan dan pengambilan tindakan berdasarkan hasil sebuah tes. Ini berarti bahwa nilai tes yang menyimpang jauh dari kompetensi siswa seperti yang dia perlihatkan dalam jawabannya akan membawa kepada penarikan kesimpulan dan tindakan yang keliru.

Ada tiga aspek utama dari sebuah tes yang sering diperiksa validitasnya, yaitu: validitas isi, validitas yang terkait kriteria, dan validitas konstruk. Penilaian validitas aspek-aspek tersebut dilakukan melalui penelitian atau studi validitas dengan prosedur yang lazim untuk memeriksa aspek-aspek tersebut dalam sebuah tes. Validitas isi dinilai dengan meneliti representasi isi tes terhadap domain atau konstruk yang diukur oleh tes tersebut. Validitas terkait kriteria (validitas ramalan dan validitas sama saat) diteliti dengan melihat korelasi skor-skor dalam sebuah tes dengan kriteria kinerja dalam bidang berkaitan. Validitas konstruk dinilai berdasarkan sejauh mana bukti-bukti yang tersedia mendemonstrasikan hubungan logis dan matematis antara konstruk dan tes yang mengukur konstruk tersebut seperti yang dihipotesakan sebelumnya.

Adapun reliabilitas tes juga sangat penting karena dia diperlukan, walaupun tidak memadai, untuk tercapainya validitas yang baik. Reliabilitas didefinisikan sebagai "the reproduceability of a set of test results under differing conditions of situations."² Sementara itu Ahmad Muhammad al-Thabib mendefinisikan

¹Messick, S., "Validity", dalam R. L. Linn, *Educational Measurement* (3rd Ed), (New York: Macmillan, 1988), hal. 14

²Untuk lebih mendalami konsep validitas, dapat dibaca Linda Crocker & James Algina, *Introduction to Classical and Modern Test Theory*, (New York: CBS College Publishing, 1986), hal. 217 - 239

³Lyman, Howard B., *Test Scores and What They Mean* (1th ed.), (Boston: (Boston: Allyn and Bacon, 1998), hal. 14

reliabilitas sebagai kesamaan atau keserupaan skor-skor siswa yang dihasilkan dari pelaksanaan tes yang lebih dari sekali.¹⁾

Dalam menaksir reliabilitas tes, ada dua prosedur yang biasa digunakan yaitu: yang pertama membutuhkan kepada dua pelaksanaan tes dan yang kedua membutuhkan kepada sekali pelaksanaan tes. Prosedur pertama terdiri dari *alternate form method* dan *test-retest method*. Prosedur yang kedua terdiri dari *split-half methods* dan *methods based on item covariances* (metode yang berdasarkan kovarian butir soal). Pembaca yang ingin mendalami prosedur penaksiran reliabilitas dapat menelaah buku rujukan seperti pada footnote tiga.

Sedikit berbeda dengan klasifikasi yang dikemukakan Muhammad Ali al-Khauili, Muhammad Abdul Khaliq Muhammad²⁾ dan demikian juga J.B. Heaton³⁾ menyebutkan insya' dapat diperiksa dengan menggunakan salah satu dari dua metode: *al-thariqah al-inthiba'iyah* atau *impression method* dan *al-thariqah al-tahliliyyah* atau *analytic method*. *Impression method*, yang dapat dikatakan sama dengan *al-thariqah al-inthiba'iyah* atau metode kesan adalah cara memeriksa jawaban insya' berdasarkan kesan pemeriksa terhadap keseluruhan jawaban insya' yang ditulis oleh murid. Dengan metode ini, nilai yang diperoleh murid cenderung ditentukan oleh faktor keberuntungan karena sangat dimungkinkan sebuah karangan (insya') menarik bagi seorang pemeriksa tapi tidak bagi pemeriksa yang lain. Kelemahan yang paling dasar dari metode ini adalah, pertama, faktor rujukannya kepada kesan pemeriksa yang sangat mudah terpengaruh oleh subjektifitas. Faktor yang kedua adalah tidak adanya upaya untuk secara detail dan sistematis meneliti bagian-bagian dari sebuah jawaban insya' seperti organisasinya, konteks, mufradat, tata bahasa dan lain-lain.

Metode yang kedua adalah *analytic method* yang oleh al-Khauili disebut dengan *al-thariqah al-tahliliyyah* atau metode analisa. Dalam pelaksanaannya metode ini bergantung pada skema

¹⁾Al-Thabib, Ahmad Muhammad, *al-Taqwim wa al-Qiyas al-Nafsiy wa al-Tarbawiy*, (Iskandariyah: al-Maktab al-Jamiy al-Hadits, t.t), hal. 113

²⁾Muhammad, Muhammad Abdul Khaliq, *Ikhtibar al-Lughat*, (Ar-Riyadh: 'Imadah Syuun al-Maktabat, 1989), hal. 202

³⁾Heaton, J.B., *Writing English Language Tests*, (New York: Longman Inc., 1990), hal. 147

pemeriksaan yang ditentukan terlebih dahulu seperti yang terdapat pada gambar 1 di depan. Pemilihan dan penekanan (bobot) pada komponen-komponen jawaban yang hendak dinilai adalah fleksibel. Guru pada pendidikan dasar misalnya dapat merevisi tabel tersebut karena dia lebih tertarik untuk fokus kepada misalnya komponen qawaid, mufradat dan imla'. Sedangkan guru pada tingkat pendidikan yang lebih tinggi dapat menggunakan skema tersebut tanpa merevisinya tetapi memberikan bobot yang lebih besar pada komponen relevansi dan uslub/kelancaran ketika menentukan nilai, atau menambahkan komponen-komponen lain seperti organisasi karangan dan sebagainya.

C. Peningkatan Pemeriksaan

Dalam beberapa dekade terakhir tes objektif seperti tes format pilihan ganda (*multiple choice*) dan variannya mendapatkan popularitas yang luas karena kepraktisan, keluasan materi yang dapat dicakup dan objektifitas pemberian skor. Namun demikian tes subjektif seperti tes insya' tidak bisa disingkirkan begitu saja walaupun memiliki kelemahan inheren dalam penentuan skor yang objektif dan konsisten. Hal yang demikian itu karena tes insya' memiliki keunggulan dalam mengukur *maharah al-kitabah* (keterampilan menulis) dengan sifat aktif-produktif yang tidak dimiliki atau tidak dapat digantikan oleh tes objektif seperti pilihan ganda. Tes pilihan ganda hanya dapat mengukur *maharah al-kitabah* terbatas pada beberapa komponen dan kemampuan dasar.

Menyadari bahwa masing-masing jenis tes memiliki kelemahan, dalam beberapa tahun terakhir mulai ada kecenderungan melengkapi tes pilihan ganda dengan jenis-jenis tes yang lain untuk mengukur sebuah kompetensi sehingga celah kelemahan sebuah tes akan ditutupi oleh kelebihan jenis yang lain dan sebaliknya. Bahkan David Thissen, professor psychometry di Universitas North Carolina, dan Howard Wainer, peneliti utama di lembaga penerbit tes terkemuka Educational Testing Service di Amerika Serikat mengantisipasi kecenderungan menggabungkan bentuk-bentuk tes yang lain, termasuk tes tulis subjektif seperti insya', dengan tes bentuk pilihan ganda akan semakin meningkat

pada masa yang akan datang.¹ Yang menjadi tantangan yang sangat serius adalah reliabilitas dan objektivitas pemeriksa yang pada gilirannya akan berdampak negatif pada validitas skor dan interpretasi hasil tes. Banyak hasil penelitian telah dipublikasi untuk menjawab tantangan ini meskipun belum secara sempurna mengatasi kelemahan-kelemahan yang ada. Untuk saat ini penulis mendiskusikan sejumlah upaya yang dapat dilakukan untuk menekan efek dari kelemahan-kelemahan tersebut ke level terendah. Dalam uraian berikut ini akan dibahas upaya-upaya tersebut.

1. Pemahaman titik lemah tes insya'

Langkah pertama dari upaya meminimalisir kekurangan-kekurangan tes insya' adalah dengan memiliki dan memperdalam pengetahuan dan pemahaman terhadap titik-titik lemah dari tes insya'. Dengan mengetahui dan menyadari hal-hal yang menjadi kekurangan tes insya', seorang pemeriksa dapat membentengi dirinya atau menghindari dari situasi yang dapat memperburuk kelemahan tes insya'. Sebagai contoh, pemeriksa harus mengupayakan jawaban insya' yang dia periksa tidak diketahui siapa penulisnya. Hal ini untuk menghilangkan pengaruh dari faktor-faktor lain pada diri murid seperti daya tarik wajah, sikap dan perilaku, latar belakang sosial ekonomi keluarga atau faktor-faktor lain yang tidak relevan dengan kualitas jawaban yang sedang diperiksa.

Berdasarkan hasil penelitian, Heaton menegaskan bahwa *unreliabilitas* adalah sifat yang tidak dapat dipisahkan dari penentuan skor jawaban insya'. Dalam menilai jawaban insya', pemeriksa sangat sulit untuk mencapai derajat reliabilitas atau konsistensi yang memadai. Seorang pemeriksa tidak saja menghadapi kesulitan untuk sampai pada skor yang sama dengan skor pemeriksa-pemeriksa lain dalam memeriksa jawaban yang sama, tetapi bahkan dia di luar kendalinya akan berlaku inkonsisten dalam memberi skor untuk jawaban yang sama pada waktu yang berbeda (*mark/re-mark reliability*). Disamping bersumber pada guru, sumber *unreliabilitas* juga bisa bercampur dengan yang

¹Thissen, David dan Wainer, Howard, *Test Scoring*, (New Jersey: Laurence Erlbaum, 2001), hal. 3

bersumber pada murid. Seorang murid yang menjawab dua tes insya' dalam tema yang serupa dan dengan tingkat kesulitan yang relatif sama tidak dapat dijamin menghasilkan jawaban dengan skor atau kualitas jawaban yang sama (*test/re-test reliability*).¹

Untuk merangkum secara garis besar titik-titik lemah tes insya' dan pemeriksaannya, pada daftar di bawah ini dapat dilihat faktor-faktor atau situasi-situasi yang berpengaruh pada kualitas pemeriksaan jawaban insya' untuk menjadi perhatian seorang pemeriksa insya'.

- a. Dua orang pemeriksa atau lebih cenderung berbeda di antara mereka dalam menilai kualitas relatif jawaban insya' seorang siswa. Pernyataan ini memperlihatkan bahwa skor insya' yang diperoleh seseorang memiliki kekurangan dalam hal nilai tunggal yang dapat disepakati oleh beberapa pemeriksa seperti halnya skor tes objektif.
- b. Dua orang pemeriksa atau lebih cenderung berbeda dalam rentang nilai, tergantung pada "kemurahan" atau "kepelitan" masing-masing. Standar yang digunakan seorang pemeriksa tentang insya' siswa yang sempurna pada suatu tingkat pendidikan cenderung berbeda dengan pemeriksa lainnya. Hal ini akan mengakibatkan seorang pemeriksa menggunakan rentang nilai yang lebar dan yang pemeriksa lain menggunakan rentang nilai yang sempit.
- c. Kesegaran dan kelelahan pemeriksa bisa mempengaruhi konsentrasi dan ketelitian pemeriksaan yang pada gilirannya akan mempengaruhi nilai yang dia berikan baik dengan penambahan nilai atau sebaliknya pengurangan.

Untuk meminimalisir unreliabilitas yang ditimbulkan oleh ketiga faktor di atas, upaya-upaya berikut dapat dilakukan oleh pemeriksa. Yang pertama, jika jawaban insya' diperiksa oleh lebih dari seorang pemeriksa dimana masing-masing menggunakan *al-thariqah al-inthiba'iyah*, maka mereka hendaklah menyepakati sebuah pedoman jawaban (sebuah versi jawaban "ideal") atau alternatif lain *scoring rubric*. *Scoring rubric* adalah sebuah pedoman penilaian yang di dalamnya dicantumkan kualitas-kualitas jawaban secara berjenjang bersama dengan nilai yang dapat

¹ Heaton, J.B., *Writing English ...*, hal. 144

diberikan terhadap setiap jawaban dengan jenjang kualitas tertentu. Dengan mempedomani *scoring rubric*, semua pemeriksa dapat diharapkan akan memberikan nilai insya' seorang siswa dengan perbedaan nilai yang lebih minimal antar pemeriksa yang satu dengan pemeriksa lainnya. Sama halnya jika insya' diperiksa oleh hanya satu orang pemeriksa, *scoring rubric* akan membantu memastikan insya' dengan kualitas lebih baik akan mendapatkan skor lebih tinggi dibanding insya' dengan kualitas yang lebih rendah.

Demikian juga halnya jika para pemeriksa menggunakan *al-thariqah al-tahliliyyah*, maka hendaklah mereka menyepakati unsur-unsur yang akan dinilai dan rentang nilai untuk masing-masing unsur sehingga dapat mempersempit perbedaan di antara mereka dalam berbagai aspek terkait pemeriksaan. Adapun jika pemeriksaan dilakukan oleh satu orang, maka penggunaan *al-thariqah al-tahliliyyah* mencegah pemeriksa dari perlakuan tidak adil dalam menilai insya'-insya' siswanya.

Yang kedua, dengan *scoring rubric* dan tabel analisa unsur-unsur yang dinilai, para pemeriksa akan terbimbing dalam pemberian skor terhadap sebuah jawaban sehingga dapat menekan perbedaan berdasarkan "kemurahan" dan "kepelitan" individual yang disebutkan di depan. Standar individu masing-masing pemeriksa tentang insya' yang sempurna atau yang sangat kurang bagi siswa pada level pendidikan tertentu yang mempengaruhi lebar rentang nilai lebur menjadi standar bersama.

Yang ketiga adalah menjadwalkan pekerjaan pemeriksaan sedemikian rupa sehingga ketika merasa lelah, pemeriksa dapat mengambil waktu untuk istirahat tetapi juga tidak terselingi oleh waktu yang terlalu lama untuk mulai kembali memeriksa jawaban yang lain yang mengakibatkan dia "lupa" dengan standar yang dia gunakan pada pemeriksaan jawaban-jawaban sebelumnya.

Dalam kaitan ini penting juga untuk diperhatikan bahwa semua murid sebaiknya ditugaskan untuk mengerjakan tugas insya' dengan tema yang sama. Jika sebagian siswa ditugaskan menulis insya' yang berbeda dengan siswa-siswa lainnya, maka dikhawatirkan keseragaman standar penilaian akan semakin sulit untuk dicapai.

¶. Unsur-unsur yang mempengaruhi penilaian

Idealnya, nilai jawaban insya' hanya didasarkan pada kualitas insya' yang mencakup komponen-komponen yang telah ditentukan untuk dipertimbangkan dalam penentuan nilai. Namun dalam kenyataannya, jawaban insya' yang hadir di depan seorang pemeriksa juga membawa kualitas-kualitas lain yang kadang-kadang tanpa disadari mempengaruhi nilai yang dia berikan. Di antara unsur-unsur yang mempengaruhi penilaian insya' adalah sebagai berikut.

a. Tulisan tangan dan tampilan umum dari tulisan murid

Mengenali bentuk-bentuk huruf Arab, kemampuan untuk menulisnya dengan baik, kebersihan dan kerapian penulisan adalah hal penting yang harus dipelajari dan dikuasai oleh murid. Kemampuan ini bahkan merupakan fondasi yang mutlak diperlukan murid untuk dapat melakukan kemampuan pada tingkat yang lebih tinggi, yaitu kemampuan insya'. Bagaimana seorang murid dapat menulis sebuah insya' jika kemampuan menulis huruf-huruf dan merangkainya untuk membentuk kata-kata dan kalimat tidak dia kuasai? Tetapi dalam penilaian insya' kemampuan ini tidaklah merepresentasikan salah satu kemampuan insya' yang diukur. Kemampuan ini termasuk dalam kemampuan dasar menulis huruf dan dapat diperhalus dengan kemampuan kaligrafi. Karena itu dalam menilai jawaban insya', seorang guru tidak perlu memasukkan unsur ini dalam penilaian. Kalau dia mendapati jawaban insya' yang masih memiliki kekurangan dalam kemampuan ini, maka selayaknya murid yang menulis jawaban tersebut "diturunkan" dulu sampai dia menguasainya secara memadai.

b. Pengetahuan terdahulu pemeriksa tentang murid

Meskipun seorang pemeriksa sangat dianjurkan untuk memeriksa sebuah jawaban insya' dengan tidak mengetahui murid mana yang menulisnya, kadang-kadang seorang guru sudah mengenali murid-muridnya dengan sangat baik. Sehingga walaupun dia tidak melihat nama murid yang menulis jawaban, dia secara serta merta dapat mengetahui murid mana yang ada dibalik jawaban tersebut dari karakter tulisan tangannya atau ciri-ciri lain yang terlihat pada jawaban tersebut. Hal ini membawa kepada

situasi dimana jawaban tidak lagi hadir sendiri tetapi bersama dengan penulisnya. Disamping pengaruh yang telah diuraikan di depan, situasi ini juga membawa pengaruh yang lain. Pemeriksa, karena sudah mengetahui siapa penulis jawaban dan mengenalnya secara sangat baik, kadang-kadang berusaha memperbaiki kesalahan atau kekurangan insya' dengan membawa keyakinan tertentu. Misalnya meyakini dalam batinnya bahwa "yang dimaksud oleh murid yang menulis jawaban adalah seperti ini" dan kemudian membubuhkan nilai sesuai koreksinya tersebut. Koreksi seperti ini harus dihindari oleh seorang pemeriksa.

3. Pengukuran sub-skill insya' dengan tes objektif

Seperti telah dikemukakan di awal tulisan ini, sebagian sub-skill kemampuan insya' seperti imla' dapat diukur dengan tes objektif. Hanya saja, sub-skill kemampuan insya' yang diukur dengan tes objektif lebih condong kepada kemampuan dalam konteks mengidentifikasi atau mengenali, sementara sub-skill kemampuan insya' yang diukur sebagaibagian dari keseluruhan kemampuan insya' lebih merupakan kemampuan aplikatif. Namun secara logis dapat diyakini bahwa sub-skill dalam dua konteks yang berbeda tersebut mempunyai korelasi kuat positif. Hal ini membuka peluang bagi guru atau pemeriksa untuk mengambil nilai sub-skill tersebut dari tes objektif atau dengan cara yang lain membandingkan dengan hasil komponen tersebut dari pemeriksaan jawaban insya' dengan ukuran yang didapatkan dari tes objektif. Perbandingan kedua ukuran tersebut kemudian dilanjutkan dengan pengolahan untuk menghasilkan sebuah ukuran atau nilai tunggal.

Dalam penilaian mufradat, ada aspek yang juga dapat diukur dengan tes objektif. Keragaman mufradat yang dikuasai siswa, misalnya, dapat diukur dengan menghitung jumlah mufradat yang berbeda yang digunakan siswa dalam merumuskan kalimat-kalimat. Skor sub-skill mufradat dari tes objektif ini selanjutnya dapat diolah bersama dengan skor mufradat dari penilaian subjektif untuk menghasilkan skor gabungan.

4. Menggabungkan dua metode pemeriksaan

Pemeriksaan insya' dengan prosedur *multi-rater* dalam *al-thariqah al-jama'iyah* merupakan sebuah upaya untuk menekan bias subjektifitas. Tetapi dalam prosedur ini bila masing-masing pemeriksa tidak secara detail memeriksa jawaban atau tidak mendasarkannya pada suatupedoman dengan deskripsi rentang kualitas yang mungkin muncul dan nilai terkait, maka bias subjektifitas bisa saja semakin besar. Oleh karena itu untuk meningkatkan kualitas pemeriksaan jawaban, *al-thariqah al-jama'iyah* sebaiknya digabungkan dengan metode yang lain. Pada uraian berikut ini akan diuraikan dua pilihan penggabungan yang dapat diambil.

Pilihan pertama yang dapat dilakukan adalah penggabungan dengan *al-thariqah al-tahliliyyah* dimana semua pemeriksa yang terlibat akan mempedomani tabel pemeriksaan unsur yang sama. Penggabungan kedua metode ini akan berpengaruh dalam dua hal. Yang pertama akan mendekatkan *sense of value* di antara para pemeriksa sebagai efek dari penggunaan *al-thariqah al-tahliliyyah*. Yang kedua, penggabungan ini secara optimal akan meminimalisir unsur subjektifitas sebagai efek dari penerapan masing-masing metode dan juga akibat dari penggabungan keduanya.

Perlu diperhatikan bahwa karena pemeriksa lebih dari satu orang maka sebuah catatan yang menjelaskan maksud detail dari unsur-unsur yang dinilai hendaklah dibuat secara spesifik. Pada unsur tata bahasa, misalnya, yang dinilai adalah sejauh mana kebenaran tata bahasa yang digunakan dalam insya' tersebut. Sementara dalam unsur kosa kata, yang menjadi fokus penilaian adalah sejauh mana kesesuaian kosa kata yang digunakan dengan makna yang dimaksudkan dan kalimat-kalimat yang digunakan.

Lebih jauh tabel *al-thariqah al-tahliliyyah* dapat dimodifikasi untuk memberikan pembobotan (*weighting*) pada aspek sub-skill insya' tertentu. Contoh modifikasi tabel tersebut dapat dilihat pada gambar di bawah ini.

Unsur	1	2	3	4	5	6	7	8	9	10	Nilai/ unsur
Tanda baca											
Qawaid											
Relevansi											
Organisasi											
Uslub/kelancaran anda baca											
Total											

Gambar. 7. Modifikasi tabel pemeriksaan insya' dengan *al-thariqah al-tahliliyyah*

Pada table di atas terlihat bahwa pemeriksa memberikan bobot lebih pada aspek relevansi, organisasi dan uslub/kelancaran dengan pertimbangan, misalnya, aspek-aspek tersebut lebih relevan dengan kemampuan insya' dibandingkan kedua aspek lainnya atau lebih sesuai dengan level sekolah siswa-siswa yang menuliskan insya' tersebut. Perlu digarisbawahi bahwa *al-thariqah al-tahliliyyah* sangat sesuai digunakan untuk sebagian besar konteks pembelajaran, karena pemeriksaan dengan metode ini akan secara jelas memperlihatkan titik-titik kelemahan siswa. Hal ini akan mempermudah guru dalam mengidentifikasi kompetensi-kompetensi yang harus diperbaiki.

Pilihan kedua adalah penggabungan dengan *al-thariqah al-intihaiyyah*. Dalam pilihan kedua ini, para pemeriksa haruslah terlebih dahulu mengembangkan dan menyepakati scoring rubric yang akan menjadi rujukan bagi mereka dalam penentuan nilai. Sebagai ilustrasi, di bawah ini ditampilkan sebuah contoh scoring

rubric yang penulis adopsi dari contoh rubric yang dikemukakan oleh Gronlund.¹

ξ – Menarik secara keseluruhan Mengalir lancar, transisi yang mulus Terorganisir dengan baik Menggunakan susunan kalimat yang baik dan tanda baca yang tepat
ϣ – Menarik pada umumnya Mengalir lancar, sebagian transisi tidak mulus Terorganisir tetapi ada beberapa kelemahan Ada sedikit kesalahan tanda baca
ϒ – Kurang menarik karena lompatan-lompatan fokus Kelancaran tertahan akibat transisi yang tidak mulus Organisasi lemah, menyimpang dari topik Ada sejumlah kesalahan tanda baca yang serius.
ϑ – Tidak ada fokus yang jelas Kacau dan campur aduk Tidak terorganisir Banyak kesalahan tanda baca dan susunan kalimat yang lemah

Gambar 3. Scoring rubric pemeriksaan insya' dengan *al-thariqah al-inthibaiyyah*

Al-thariqah al-inthibaiyyah lebih tepat diterapkan pada level pendidikan yang lebih tinggi, dimana masalah mufradat, qa'idah dan imla' tidak lagi menjadi perhatian. Penulis insya' dianggap sudah menguasai komponen-komponen tersebut. Karena yang menjadi sasaran penilaian lebih pada gaya bahasa, organisasi dan kemampuan insya' menciptakan kesan tertentu pada diri pembaca.

D. Penutup

Menghasilkan skor dengan kualitas yang memenuhi kriteria objektif, reliabel dan valid mulai dari tahap menulis tes hingga

¹Gronlund, Norman E., *Assessment of Student Achievement*, (Boston: Allyn and Bacon, 1998), hal. 106

tahap memeriksa jawaban tes haruslah menjadi komitmen setiap penulis tes dan pemeriksa jawaban. Kualitas seperti yang disebutkan di atas lebih memungkinkan untuk dicapai dalam tes objektif dibandingkan tes subjektif seperti *insya'*. Namun demikian tidak berarti ancaman subjektifitas, unreliabilitas dan invaliditas tidak dapat diminimalisir. Ada sejumlah upaya yang dapat dilakukan untuk meningkatkan objektifitas, reliabilitas dan validitas skor *insya'*. Upaya-upaya yang dapat dilakukan antara lain adalah dengan memahami titik lemah tes *insya'* dan faktor-faktor tidak relevan yang dapat mempengaruhi skor. Dengan memahami kelemahan tes *insya'* dan faktor yang berpengaruh seorang pemeriksa mempertimbangkan solusinya dan menghindari keterpaparan terhadap faktor-faktor tersebut.

Upaya-upaya lain yang dapat dilakukan adalah dengan mengkombinasikan metode pemeriksaan *al-thariqah al-jamaiyyah* dengan *al-thariqah al-tahliliyyah* dan melibatkan ukuran sub-skill *insya'* yang diambil dari tes objektif. Pada level pendidikan yang lebih tinggi, *al-thariqah al-inthibaiyyah* lebih sesuai untuk diterapkan baik sebagai metode tunggal atau dengan digabungkan dengan *al-thariqah al-jamaiyyah*. Upaya peningkatan hasil pemeriksaan juga dapat dilakukan dengan menggunakan statistik sederhana untuk mengukur perbendaharaan atau keragaman mufradat yang digunakan siswa dalam *insya'* yang dia tulis.

Daftar Pustaka

Al-Khauili, Muhammad Ali, *Al-Ikhtibarat al-Lughawiyah*, Al-Urdun: Dar al-Falah li al-Nasyr wa al-Tauzi', ٢٠٠٠

Al-Thabib, Ahmad Muhammad, *al-Taqwim wa al-Qiyas al-Nafsiy wa al-Tarbawiy*, Iskandariyah: al-Maktab al-Jamiyy al-Hadits, t.t

Crocker, Linda & Algina, James, *Introduction to Classical and Modern Test Theory*, New York: CBS College Publishing, ١٩٨٦

Gronlund, Norman E., *Assessment of Student Achievement*, Boston: Allyn and Bacon, ١٩٩٨

Heaton, J.B., *Writing English Language Tests*, New York: Longman Inc, ١٩٩٥

Linn, R. L. *Educational Measurement* (٣rd Ed), New York: Macmillan, ١٩٨٨

Lyman, Howard B., *Test Scores and What They Mean* (٦th ed.), Boston: Allyn and Bacon, ١٩٩٨

Matsna HS, Moh. & Mahyudin, Erta, *Pengembangan Evaluasi dan Tes Bahasa Arab*, Tangerang Selatan: Alkitabah, ٢٠١٢

Muhammad, Muhammad Abdul Khaliq, *Ikhtibarat al-Lughat*, Ar-Riyadh: 'Imadah Syuun al-Maktabat, ١٩٨٩

Thissen, David & Wainer, Howard, *Test Scoring*, New Jersey: Laurence Erlbaum, ٢٠٠١