

**PENGEMBANGAN INSTRUMEN
PENILAIAN UNJUK KERJA PADA PEMBELAJARAN DENGAN MODEL
PROBLEM SOLVING BERBASIS TIK**

**Ibnu Wachyudi*, YLSukestiyarno,
St. Budi Waluya.**

Email :ibnuw07@yahoo.co.id

ABSTRAK

Penelitian ini merupakan penelitian dan pengembangan (*Research and Development*) bertujuan untuk mengembangkan instrumen penilaian unjuk kerja pada pembelajaran matematika materi segitiga kelas VII dengan model *problem solving* berbasis TIK yang valid dan reliabel. Model pengembangan instrument penelitian ini mengacu pada model *Borg and Gall* (1987), yang disesuaikan. Instrumen ini dapat mengukur tiga ranah pada unjuk kerja siswa. Ranah afektif pada karakter tanggung jawab, ranah psikomotor pada keterampilan memecahkan masalah dan ranah kognitifnya ditujukan pada siswa agar mampu memecahkan masalah. Instrumen ini diterapkan pada model pembelajaran *problem solving* yang mempunyai empat langkah yaitu: memahami, merencana, melaksanakan dan mereview. Setiap masalah dipecahkan dengan empat langkah dan penyelesaian tugas rumahnya melalui TIK khususnya fasilitas e-mail. Produk instrumen penilaian unjuk kerja ini divalidasi isi oleh 7 ahli dan praktisi terdiri dari : bidang pembelajaran dan materi, evaluasi, serta pengguna. Instrumen yang dikembangkan diujicobakan kepada 50 siswa SMP Muhammadiyah 2 dan 95 siswa SMP Negeri 5 Kota Tegal. Validitas instrument penilaian unjuk kerja pada karakter tanggung jawab dan keterampilan memecahkan masalah diukur menggunakan *analisa factor* eksploratori berbantuan *software SPSS versi 16* dengan hasil semua butir amatan nilai *loading factornya* melampaui *cut of poin* (0,55). Validitas intrumen penilaian unjuk kerja pada kemampuan memecahkan masalah diukur menggunakan rumus *product moment*, di samping dilakukan uji daya beda dan uji tingkat kesukaran. Hasil uji ini menghasilkan 6 butir soal valid dari 8 butir soal yang mewakili 4 indikator. Reliabilitas instrumen penilaian unjuk kerja pada karakter tanggung jawab dan keterampilan memecahkan masalah diukur menggunakan *Intraclass Correlation Coefisient (ICC)*, menghasilkan nilai koefisien yang melampaui batas minimal (0,70). Reliabelitas instrument penilaian unjuk kerja pada kemampuan memecahkan masalah diukur menggunakan rumus *Sperman - Brown*, menghasilkan nilai reliabilitas sebesar 0,810. Keduanya berbantuan *software SPSS* versi 16.

Kata Kunci : Pengembangan Instrumen, Penilaian Unjuk Kerja, *ICC*.

Pendahuluan

Undang – undang Nomor 14 Tahun 2005 tentang Guru dan Dosen pasal 20 salah satu butirnya menyatakan dalam melaksanakan tugas keprofesionalannya guru mempunyai kewajiban merencanakan pembelajaran, melaksanakan proses pembelajaran yang bermutu, serta menilai dan mengevaluasi hasil pembelajaran. Kegiatan menilai yang dilakukan oleh guru mempunyai makna sebagai suatu usaha mengumpulkan informasi tentang hasil belajar. Penilaian yang dilakukan guru merupakan bentuk tanggung jawab yang harus dilaksanakan guru setelah atau selama proses pembelajaran. Hasil penilaian ini akan digunakan sebagai bahan dalam kegiatan evaluasi yang selanjutnya akan digunakan untuk mengambil keputusan mengenai hasil pembelajarannya (Tucker et al dalam Gareis, 2007:20).

Penilaian dalam pembelajaran mempunyai tujuan diantaranya adalah (1) umpan balik bagi peserta didik untuk mengetahui kelemahan dan kelebihan dalam mengikuti pembelajaran

serta hasil usahanya, (2) umpan balik bagi guru untuk mengetahui efektifitas kegiatan pembelajaran yang dilakukannya, (3) informasi bagi orang tua sekaligus sebagai bentuk pertanggungjawaban sekolah dalam mengelola kegiatan pembelajaran, (4) sebagai pertimbangan dalam memberikan penghargaan dan motivasi kepada peserta didik agar meningkatkan usaha belajarnya (Wardhani, 2004: 1)

Penilaian pembelajaran perlu dilakukan secara menyeluruh dan berkesinambungan, hal ini selaras dengan pernyataan Allen dalam Praslova (2010: 1) bahwa penilaian merupakan suatu proses yang berkelanjutan yang dirancang untuk memantau proses pembelajaran dan meningkatkan hasil belajar. Untuk mencapai tujuan penilaian dapat dilakukan berbagai cara, yang pada akhirnya adalah mampu “memotret” peserta didik dalam kegiatan belajarnya. Pemotretan terhadap peserta didik perlu dilakukan secara menyeluruh baik pada prestasi akademik (kognitif), keterampilan (psikomotor), tingkah laku dan sikapnya (afektif) agar hasil “pemotretan” lebih dekat dengan kondisi kemampuan nyata peserta didik yang “dipotret”. Kegiatan “pemotretan” atau penilaian hasil belajar ini dapat dilakukan dengan dua cara yakni Tehnik tes dan non tes (Sukardi, 2009: 9 – 11).

Penilaian hasil belajar merupakan penilaian terhadap kemampuan seseorang dalam menyelesaikan suatu kegiatan yang hasilnya menyangkut aspek kognitif, afektif dan psikomotor (Sukestiyarno, 2012: 11). Hal ini sejalan dengan Hidayat dan Maryani (Masrukan, 2013: 33) bahwa dalam melakukan penilaian unjuk kerja peserta didik harus menunjukkan beberapa unsur sebagai berikut: (1) *Knowledge* atau pengetahuan (2) *Reasoning* yang berarti penalaran atau aplikasi pengetahuan dalam konteks pemecahan masalah (3) *Skill* yaitu kecakapan siswa dalam bertanya, keterampilan berkomunikasi, karya, visual dll (4) *Product* yaitu berbagai macam karya siswa dan (5) *Affect* yaitu menggambarkan tentang tingkah laku, minat, nilai motivasi dan konsep diri. Penilaian unjuk kerja tersebut mempunyai unsur secara garis besar terdiri dari aspek kognitif, psikomotor dan afektif yang mempunyai kesamaan maksud dengan penilaian hasil belajar.

Penilaian hasil belajar atau penilaian unjuk kerja diharapkan mampu menilai kompetensi peserta didik secara menyeluruh. Kompetensi lulusan yang mencakup aspek afektif, kognitif dan psikomotor ini harus dikembangkan melalui proses pembelajaran maupun penilaian.

Pada lima belas tahun terakhir banyak ahli pendidikan mulai melakukan penelitian mengenai peran pentingnya ranah afektif khususnya pada pembelajaran matematika (Drodge dan Red dalam Peter, 2006: 193), namun di negara kita sekolah - sekolah umumnya terfokus pada kegiatan penilaian yang mengarah pada ranah kognitif (Wardani, 2004: 1) dan kurang memperhatikan pada ranah afektif (tingkah laku dan sikap). Semestinya antara ranah kognitif dan afektif berimbang, sebab pada pembelajaran mata pelajaran matematika afektif dan kognitif dalam prakteknya saling pengaruh mempengaruhi (Margaret & Brown, 2012: 186).

Permendikbud No. 66 Tahun 2013, menyatakan Ujian Nasional merupakan kegiatan pengukuran kompetensi tertentu yang dicapai peserta didik dalam rangka menilai pencapaian Standar Nasional Pendidikan yang dilaksanakan secara nasional yang dalam pengukurannya menggunakan tehnik tes berbentuk pilihan ganda, pemilihan bentuk tes ini karena kepraktisannya dan lebih mudah dalam mengoreksi. Secara tidak langsung pemilihan bentuk tes ini berpengaruh pada sekolah – sekolah di Indonesia, tak terkecuali di Kota Tegal empat dari lima guru Sekolah Menengah Pertama (SMP) khususnya guru mata pelajaran matematika dalam penilaiannya menggunakan ranah kognitif dan melupakan dua ranah penilaian lainnya, ranah afektif dan psikomotor.

Upaya perubahan penilaian yang terfokus pada hanya aspek kognitif telah dilakukan pemerintah melalui kegiatan pelatihan atau menerbitkan buku pedoman pengembangan penilaian ranah afektif maupun psikomotor, meski belum menampakkan hasil maksimal. Upaya yang belum menampakkan hasil maksimal ini dimungkinkan karena tidak sampainya buku pedoman pengembangan instrumen penilaian ranah afektif atau psikomotor di tingkat satuan pendidikan secara luas di samping kemampuan sumber daya manusianya yang terbatas dalam mengembangkan instrumen penilaian unjuk kerja ranah afektif maupun psikomotor.

Selama ini berdasarkan dari hasil pengamatan, guru melakukan penilaiannya terfokus hanya mengukur penguasaan pengetahuan (ranah kognitif) saja. Peneliti meyakinkan hasil

pengamatannya melalui wawancara terhadap lima guru matematika SMP di Kota Tegal, yang diambil secara *purposive sampling*. SMP Negeri 5 mewakili sekolah negeri yang belum bertaraf SSN. SMP Negeri 17 dan 19 mewakili sekolah yang bertaraf SSN yang melaksanakan kurikulum 2013. Perwakilan sekolah swasta adalah SMP Muhammadiyah 1 dan SMP Atmaja Wacana. Wawancara dilakukan untuk mendapatkan informasi awal dilakukan pada tanggal 8 sampai dengan 9 Oktober 2013, termasuk mewawancari kepala sekolahnya kecuali kepala SMP Negeri 17 yang sedang menjalankan ibadah haji.

Hasil wawancara tiga dari enam orang guru menyatakan penilaian yang dilakukan lebih sering mengukur kemampuan kognitif namun diakui juga pernah melakukan penilaian ranah afektif secara tak terencana dan tanpa acuan yang jelas, sehingga subjektivitas penilaiannya sangat besar. Lebih mengejutkan, guru yang dimintai informasi mengenai cara penilaian ranah afektifnya bahkan bertanya balik mengenai pengertian penilaian ranah afektif, karena guru merasa lupa. Salah seorang dari tiga orang guru yang melakukan penilaian afektif menyatakan bahwa penilaian afektif yang dilakukan hanya jika dipandang butuh karena nilai peserta didik tidak mencapai KKM, sehingga peserta didik yang teramat patuh, sopan dan *manut* maka nilainya dinaikkan mencapai KKM.

Kenyataan di lapangan lainnya, guru yang telah melakukan penilaian afektif atau psikomotor kesulitan dalam membubuhkan nilai pada buku daftar nilainya, karena buku daftar nilai yang disediakan sekolah hanya menyediakan kolom penilaian kognitif. Hal ini dibenarkan oleh tiga dari empat orang kepala sekolah bahwa buku nilai hanya tersedia kolom penilaian kognitif saja. Kenyataan ini menunjukkan terabaikannya penilaian afektif juga psikomotor peserta didik. Berbeda dengan sekolah yang telah melaksanakan kurikulum 2013 pada buku daftar nilainya telah disediakan kolom penilaian afektif namun belum dilengkapi kolom penilaian psikomotor.

Terfokusnya penilaian pada ranah penguasaan pengetahuan (kognitif) berimplikasi pada penilaian proses pembelajaran yang dilaksanakan oleh guru di dalam kelas terabaikan. Hal ini disebabkan dalam proses pembelajaran tersebut guru tidak menyadari apa yang dilaksanakan dalam kelas sebagai proses afektif dan psikomotor, semestinya penting disadari karena untuk perbaikan proses pembelajaran yang telah dilakukannya (Stanislaw et al, 2012: 216).

Pembelajaran yang dilakukan guru menjadi membosankan dan memunculkan rasa ketidaksukaan terhadap mata pelajaran matematika, karena matematika dirasakannya sebagai pelajaran yang sulit dan menumbuhkan rasa tidak percaya diri (Margaret & Brown, 2012: 186). Pembelajaran yang menyertakan penilaian proses selama pembelajarannya akan diketahui bagaimana keaktifan dan keterampilan peserta didik selama pembelajaran berlangsung. Dan ini akan melunturkan rasa bosan, tidak suka atau rasa malas terhadap pembelajaran matematika.

Terkait dengan permasalahan penilaian, Suyanto dalam Mulyani (2012: 265) mengatakan bahwa sampai saat ini pendidikan di semua jenjang dalam pelaksanaan penilaian masih lebih mementingkan aspek kognitif, sedangkan aspek afektif masih ditelantarkan. Untuk mengatasi permasalahan yang terkait dengan penilaian seperti telah dikemukakan di atas, sangatlah penting untuk dikembangkan instrumen penilaian komprehensif yang mengukur ranah afektif, psikomotor dan kognitif selama dan setelah proses pembelajaran.

Pada penelitian ini mengambil materi segitiga kelas 7 dengan pertimbangan bahwa materi segitiga merupakan materi *esensial* yang banyak dikaitkan dengan permasalahan kehidupan sehari – hari. Pembelajaran yang dilakukan mengikuti kurikulum yang berlaku yakni mengintegrasikan pendidikan karakter dan TIK pada mata pelajaran matematika. Hal ini juga mengiringi perubahan kurikulum 2006 menjadi kurikulum 2013 yang meletakkan semua kompetensi secara bersama – sama baik kompetensi keaktifan, keterampilan dan pengetahuan sebagai kompetensi lulusan (Tim Sergur Unnes, 2013a: 10 dan Kemendikbud, 2013: 79).

Unjuk kerja pada aspek afektif yang akan dikembangkan adalah karakter tanggung jawab, aspek psikomotor yang akan dikembangkan adalah keterampilan pemecahan masalah dan pada aspek kognitif yang akan dikembangkan adalah kemampuan pemecahan masalah. Sehubungan dengan hal tersebut perlu adanya suatu pendekatan atau model pembelajaran tertentu yang dapat meningkatkan kemampuan peserta didik dan hasil belajar peserta didik. Pada penelitian ini akan diterapkan model *problem solving* yang mempunyai langkah menyelesaikan masalah yang terkait dengan materi dan kehidupan sehari – hari.

Pembelajaran ini pada prinsipnya sebagai sarana mengembangkan instrumen unjuk kerja,

Model pembelajaran *problem solving* membelajarkan matematika yang terpusat pada peserta didik sehingga terjadi saling ketergantungan positif di antara peserta didik untuk mencapai tujuan pembelajaran. Kegiatan belajar yang berpusat pada peserta didik bisa dalam bentuk diskusi, mengerjakan tugas bersama, saling membantu dan saling mendukung dalam memecahkan masalah. Melalui kegiatan pembelajaran yang efektif, peserta didik lebih termotivasi, percaya diri, mampu membangun hubungan antar peserta didik serta mampu menggunakan strategi berpikir tingkat tinggi.

Model ini menuntut para peserta didik memiliki kemampuan yang baik dalam menyadari adanya masalah yang akan dipecahkan dengan cara mengidentifikasi masalah tersebut. Penggunaan pengalaman sebelumnya atau informasi akan bermanfaat untuk penyusunan hipotesis pemecahan masalah. Pengujian hipotesis bisa dilakukan berulang kali untuk mendapat solusi yang mungkin. Dan terakhir cek terhadap solusi dan menyusun kesimpulan berdasarkan bukti yang ada.

Selama proses pembelajaran dilakukan penilaian unjuk kerja meliputi karakter tanggung jawab (afektif), keterampilan pemecahan masalah (psikomotor) dan kemampuan pemecahan masalah (kognitif). Instrumen penilaian karakter tanggung jawab didesain agar dapat mengukur rasa tanggung jawab siswa dalam usaha memecahkan masalah menurut polya yang terdiri dari empat langkah, yaitu: memahami, merencana, melaksanakan dan mereview. Instrumen penilaian keterampilan pun demikian didesain agar dapat mengukur keterampilan siswa dalam upaya memecahkan masalah menurut langkah pemecahan masalah Polya. Instrumen penilaian kemampuan masalah juga menggunakan langkah – langkah penyelesaian pemecahan masalah polya.

Instrumen penilaian unjuk kerja tersebut akan dikaji secara mendalam agar meniadakan derajat instrumen yang valid dan reliabel. Azwar (2011: 2) menyatakan instrumen yang baik adalah instrumen yang mampu memberikan informasi yang dapat dipercaya. Alat ukur atau instrumen mempunyai kriteria tertentu sehingga dapat disebut sebagai instrumen yang baik atau layak. Kriteria tersebut adalah instrumen harus valid dan reliabel.

Instrumen penilaian ini dikembangkan dengan harapan akan memberikan hasil penilaian terhadap peserta didik selama dan setelah mengikuti proses pembelajaran. Dengan pengembangan instrumen ini juga diharapkan dapat melakukan penilaian secara komprehensif dan dapat meningkatkan kualitas pembelajaran. Peningkatan kualitas pembelajaran akan dapat meningkatkan prestasi belajar peserta didik. Peningkatan prestasi belajar secara luas akan memberikan andil dalam peningkatan kualitas pendidikan secara nasional.

Terkait dengan latar belakang tersebut, pada penelitian ini akan dikembangkan instrument penilaian unjuk kerja pada pembelajaran matematika dengan menggunakan model *Problem solving* berbasis TIK yang valid dan reliable. Instrumen unjuk kerja ini meliputi ranah afektif pada karakter tanggung jawab, ranah psikomotor pada keterampilan memecahkan masalah dan ranah kognitifnya pada kemampuan memecahkan masalah.

Metode Penelitian

Metode penelitian diartikan sebagai cara ilmiah untuk mendapatkan data dengan tujuan dan kegunaan tertentu (Sugiyono, 2009: 3). Penelitian ini menggunakan gabungan metode kuantitatif dan kualitatif secara bersamaan untuk memperoleh analisis komprehensif atas masalah penelitian sebagaimana yang dikemukakan Creswell (2013: 23) atau biasa dikenal dengan istilah *concurrentmix method* (metode campuran konkurent).

Model penelitian pengembangan yang relevan dengan penelitian ini yang tentunya juga mempertimbangkan karakteristik produk dan efisiensi proses pengembangan model instrumen penilaian unjuk kerja ini, menggunakan penelitian *Research and Development* (R&D) Borg and Gall terbatas sampai pada langkah kesembilan. Pada langkah kesembilan ini pengembangan instrumen dengan hasil akhir produk instrumen yang valid dan reliabel telah tercapai.

Kesembilan langkah tersebut adalah: (1) *Research and Information Collection*, (2) *Planning*, (3) *Development of the preliminary form of product*, (4) *Preliminary field testing*, (5) *Main product revision*, (6) *Main field tes*, (7) *Operasional product revision*, (8) *Operational*

field testing, (9) Final Product Final Revision.

Untuk memperoleh data yang terpercaya diperlukan instrumen yang valid dan reliabel. Guna memenuhi hal tersebut, produk instrumen penilaian unjuk kerja yang dikembangkan ini divalidasi isi oleh 7 ahli dan praktisi terdiri dari : bidang pembelajaran dan materi, evaluasi, serta pengguna. Instrumen yang dikembangkan diujicobakan sebanyak dua kali, yakni kepada 50 siswa SMP Muhammadiyah 2 sebagai uji coba skala terbatas dan 95 siswa SMP Negeri 5 Kota Tegal sebagai uji coba skala luas.

Validitas instrument penilaian unjuk kerja pada karakter tanggung jawab dan keterampilan memecahkan masalah diukur menggunakan *analisa factor* eksploratori berbantuan *software SPSS versi 16*. Validitas intrumen penilaian unjuk kerja pada kemampuan memecahkan masalah diukur menggunakan rumus *product moment*, di samping dilakukan uji daya beda dan uji tingkat kesukaran.

Reliabilitas instrumen penilaian unjuk kerja pada karakter tanggung jawab dan keterampilan memecahkan masalah diukur menggunakan *Intraclass Correlation Coefisient (ICC)*. Reliabelitas instrument penilaian unjuk kerja pada kemampuan memecahkan masalah diukur menggunakan rumus *Sperman - Brown*, keduanya dihitung dengan bantuan *software SPSS* versi 16.

Hasil dan Pembahasan

Instrumen penilaian unjuk kerja adalah alat yang dapat digunakan untuk mengetahui pencapaian peserta didik berdasarkan tindakan atau kinerja dari apa yang menjadi harapan-harapan pendidik. Menurut Trespeces yang dikutip oleh Noviyanti (2010) mengatakan bahwa penilaian kinerja (*Performance Assessment*) adalah berbagai macam tugas dan situasi dimana peserta tes diminta untuk mendemonstrasikan pemahaman dan pengaplikasian pengetahuan yang mendalam, serta keterampilan di dalam berbagai macam konteks.

Berdasar pada pemahaman yang dikemukakan di atas maka pengembangan instrumen penilaian unjuk kerja ini terbagi menjadi tiga macam, yakni (1) pada aspek afektif yang dikembangkan adalah karakter tanggung jawab, (2) pada aspek psikomotor yang dikembangkan adalah keterampilan memecahkan masalah, dan (3) pada aspek kognitif yang dikembangkan adalah kemampuan memecahkan masalah.

1.1. Instrumen Penilaian Unjuk Kerja Pada Karakter Tanggung Jawab

Puskur (2010: 10) mendiskripsikan tanggung jawab adalah sikap dan perilaku seseorang untuk melaksanakan tugas dan kewajiban yang seharusnya dilakukan terhadap diri sendiri, masyarakat, lingkungan, serta Negara dan Tuhan yang Maha Esa. Berangkat dari pemahaman ini instrumen penilaian unjuk kerja pada karakter tanggung jawab ini dikembangkan. Hasil instrumen yang dikembangkan ini merupakan instrumen penilaian pengamatan.

Azwar (2011: 2) menyatakan instrumen penilaian yang baik adalah instrumen penilaian yang mampu memberikan informasi yang dapat dipercaya. Instrumen penilaian yang baik ini mempunyai kriteria yaitu valid, reliabel, standar, ekonomis dan praktis. Sifat valid dan reliabel harus ditentukan berdasar hasil empiris lapangan sehingga diperlukan uji coba instrumen. Sifat valid dan reliabel diperlihatkan dari tingginya nilai koefisien reliabelias dan validitas hasil ukur instrumen penilaian.

Salah satu pendekatan untuk menentukan validitas konstruk menggunakan *analisa factor* (Azwar, 2011: 132) pada uji coba skala kecil menghasilkan tiga faktor yang dibentuk dari 18 butir amatan. Nilai *loading factor* butir amatan terhadap faktor 1 berturut – turut adalah 0.878, 0.867, 0.804, 0.743, 0.801, 0.811, 0.773, 0.742, 0.816, 0.634, dan 0.739. Nilai *loading factor* butir amatan terhadap faktor 2 berturut – turut adalah 0.781, 0.832, 0.901, 0.705, dan 0.798. Sedangkan nilai *loading factor* butir amatan terhadap faktor 3 adalah 0.866 dan 0.840. Nilai – nilai tersebut termasuk dalam kategori tinggi kecuali nilai 0.634 pada faktor 1. Nilai – nilai koefisien tersebut secara keseluruhan dapat menunjukkan bahwa butir – butir amatan yang dikembangkan memenuhi sifat valid karena semua nilai *loading factor* di atas nilai pembatas (*cut off point*) (Santoso, 2003: 120) dan (Dwipurwani, 2009: 3).

Pada uji skala luas *analisa factor* mereduksi 18 butir amatan (variable) menjadi tiga faktor seperti pada uji skala kecil. Nilai *loading factor* butir amatan (X_{12} , X_{13} , X_{14} , X_{15} , X_{16} , X_{17} ,

dan X_{18}) terhadap faktor 1 yang terbentuk berturut – turut adalah 0.684, 0.725, 0.800, 0.762, 0.789, 0.904, dan 0.914. Nilai *loading factor* butir amatan ($X_2, X_3, X_6, X_9, X_{10}$, dan X_{11}) terhadap faktor 2 berturut – turut adalah 0.760, 0.850, 0.849, 0.528, 0.751, dan 0.765. Sedangkan nilai *loading factor* butir amatan lainnya (X_1, X_4, X_5, X_7 , dan X_8) terhadap faktor 3 adalah 0.659, 0.758, 0.857, 0.810, dan 0.731. Semua nilai *loading factor* sebagai nilai yang menunjukkan besar korelasi antara butir amatan terhadap faktor berada di atas nilai pembatas (*cut off point*) yang ditetapkan yakni $> 0,3$ menurut Dwipurwani (2009: 3) dan Santoso (2003: 120) menetapkan *cut off point* - nya $> 0,55$, dengan demikian instrumen penilaian karakter tanggung jawab memenuhi sifat valid.

Widhiarso (2014: 15) dan Widhiarso (2014a: 2) menyatakan metode mengestimasi reliabelitas terhadap instrumen penilaian yang dilakukan oleh lebih dari dua *rater* adalah dengan menggunakan reliabelitas antar *rater* yang dihitung dengan menggunakan koefisien korelasi antar kelas (*Intraclass Correlation Coefisients, ICC*). ICC ini menunjukkan perbandingan variasi yang diakibatkan atribut yang diukur dengan variasi pengukuran secara keseluruhan.

Pada pelaksanaan uji coba skala kecil instrumen penilaian karakter tanggung jawab menghasilkan nilai koefisien untuk pasangan *rater* 1 dan *rater* 2 sebesar 0,649, pasangan *rater* 1 dan *rater* 3 nilai koefisiennya 0,665 sedangkan pasang *rater* 2 dan *rater* 3 adalah 0,842. Kriteria yang diharapkan $\bar{r} > 0,70$ (Utomo, 2012: 216) dengan demikian nilai koefisien ICC yang dihasilkan pada uji skala kecil pada instrumen penilaian karakter belum memenuhi harapan kecuali pada pasangan *rater* 2 dan *rater* 3. Hasil yang belum sesuai dengan harapan maka peneliti melakukan pencermatan terhadap skor butir – butir pengamatan yang mempunyai selisih ekstrim. Skor – skor yang mempunyai selisih ekstrim adalah pada butir pengamatan 1, 2, 5, 7, 8, 9, 11, 12, dan 13 kemudian butir – butir tersebut didiskusikan bersama para *rater* untuk dicari penyebab dari perbedaan penskoran yang berbeda ekstrim. Setelah diketahui penyebab selisih ekstrim peneliti melakukan perbaikan terhadap instrumen dan menyepakati maksud yang diamati pada butir yang menimbulkan keraguan *rater* dalam memberi skor.

Tabel 1. Peningkatan Koefisien Reliabelitas Instrumen Pengamatan Karakter Tanggung jawab

Measures	<i>Interclass Correlation Coeffisien (ICC)</i>			
	Uji Coba Skala Kecil		Uji Coba Skala Luas	
	<i>Consistentensy</i>	<i>Absolute agreement</i>	<i>Consistentensy</i>	<i>Absolute agreement</i>
<i>Single Measure</i>	0,563	0,379	0,849	0,805
<i>Average Measures</i>	0.794	0,647	0,944	0,925

Keterangan (Utomo, 2012: 348):

1. Kriteria dengan tipe uji *consistency* minimal 0,70 ($\bar{r} > 0,70$)
2. Kriteria dengan tipe uji *Absolute agreement* minimal 0,60 ($\bar{r} > 0,60$)

Hasil uji coba skala luas instrumen penilaian karakter tanggung jawab menunjukkan nilai koefisien korelasi antar kelas (*ICC*) tipe *consistentensy definition* dan *absolute agreement definition* meningkat untuk semua pasangan *rater*. Hasil uji reliabelitas dengan *ICC* dideskripsikan pada Tabel 1.

Data tersebut menunjukkan bahwa pada saat uji skala kecil koefisien reliabelitasnya (*Single Measure*) instrumen penilaian karakter tanggung jawab hanya sebesar 0,563 menggunakan analisis *ICC* tipe *Consistentensy definition*, sedangkan menggunakan analisis *ICC* tipe *Absolute agreement definition* hanya sebesar 0,379. Namun demikian setelah dilakukan revisi dan kesepakatan yang dimaksud dalam pengamatan, koefisien meningkat menjadi 0,805 dalam tipe *Absolute agreement definition* dan 0,849 dalam tipe *Consistentensy definition*. Nilai koefisien reliabelitasnya tidak hanya meningkat namun juga melampaui dari batas yang ditetapkan (Utomo, 2012: 348), dengan demikian sifat reliabel pada instrumen penilaian unjuk kerja pada karakter tanggung jawab dipenuhi.

1.2. Instrumen Penilaian Unjuk Kerja Pada Keterampilan Memecahkan Masalah

Mengkombinasikan teori pemecahan masalah Polya dalam Wardhani (2010: 35) dan Permendiknas 54 tahun 2013 mengenai keterampilan memecahkan masalah maka pengembangan instrumen penilaian unjuk kerja pada keterampilan memecahkan masalah didesain sebagai suatu keterampilan pikir dan tindakan yang efektif dan kreatif sesuai dengan yang dipelajari di sekolah maupun sumber lain dalam usaha mencari jalan keluar dari kesulitan untuk mencapai suatu tujuan. Hasil instrumen yang dikembangkan ini merupakan instrumen penilaian pengamatan.

Azwar (2011: 2) menyatakan instrumen penilaian yang baik adalah instrumen penilaian yang mampu memberikan informasi yang dapat dipercaya. Instrumen penilaian yang baik ini mempunyai kriteria yaitu valid, reliabel, standar, ekonomis dan praktis. Sifat valid dan reliabel harus ditentukan berdasar hasil empiris lapangan sehingga diperlukan uji coba instrumen. Sifat valid dan reliabel diperlihatkan dari tingginya nilai koefisien reliabilitas dan validitas hasil ukur instrumen penilaian.

Salah satu pendekatan untuk menentukan validitas konstruk menggunakan *analisa factor* (Azwar, 2011: 132) oleh karena itu pada uji coba skala kecil ini dilakukan analisa dengan menggunakan *analisa factor*. Hasil analisisnya membentuk sebuah faktor dari 16 butir amatan. Nilai *loading factor* butir amatan terhadap faktor berturut – turut adalah 0.904, 0.841, 0.896, 0.912, 0.806, 0.898, 0.937, 0.888, 0.858, 0.896, 0.923, 0.869, 0.899, 0.869, 0.876, dan 0.884. Nilai – nilai koefisien tersebut secara keseluruhan dapat menunjukkan bahwa butir – butir amatan yang dikembangkan memenuhi sifat valid karena semua nilai *loading factor* di atas nilai pembatas (*cut off point*) yakni lebih besar dari 0,55 (Santoso, 2003: 120) dan sedikit beda dengan Dwipurwani (2009: 3) menyatakan *cut off point* lebih besar 0,30.

Pada uji skala luas *analisa factor* mereduksi 16 butir amatan (variable) menjadi sebuah faktor seperti pada uji skala kecil. Nilai *loading factor* butir amatan ($x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15},$ dan x_{16}) terhadap faktor yang terbentuk berturut – turut adalah 0.909, 0.846, 0.901, 0.913, 0.804, 0.898, 0.939, 0.892, 0.862, 0.901, 0.926, 0.868, 0.898, 0.869, 0.877, dan 0.889. Semua nilai *loading factor* sebagai nilai yang menunjukkan besar korelasi antara butir amatan terhadap faktor berada di atas nilai pembatas (*cut off point*) yang ditetapkan yakni $> 0,3$ menurut Dwipurwani (2009: 3) dan Santoso (2003: 120) menetapkan *cut off point* nya $> 0,55$, dengan demikian instrumen penilaian karakter tanggung jawab memenuhi sifat valid.

Widhiarso (2011: 15) dan Widhiarso (2014a: 2) menyatakan metode mengestimasi reliabilitas terhadap instrumen penilaian yang dilakukan oleh lebih dari dua rater adalah dengan menggunakan reliabilitas antar rater yang dihitung dengan menggunakan koefisien korelasi antar kelas (*Intraclass Correlation Coefficients, ICC*). ICC ini menunjukkan perbandingan variasi yang diakibatkan atribut yang diukur dengan variasi pengukuran secara keseluruhan.

Pada pelaksanaan uji coba skala kecil instrumen penilaian karakter tanggung jawab menghasilkan nilai koefisien untuk pasangan rater 1 dan rater 2 sebesar 0,649, pasangan rater 1 dan rater 3 nilai koefisiennya 0,665 sedangkan pasangan rater 2 dan rater 3 adalah 0,842. Kriteria yang diharapkan $\bar{r} > 0,70$ (Utomo, 2012: 216) dengan demikian nilai koefisien ICC yang dihasilkan pada uji skala kecil pada instrumen penilaian karakter belum memenuhi harapan kecuali pada pasangan rater 2 dan rater 3. Hasil yang belum sesuai dengan harapan maka peneliti melakukan pencermatan terhadap skor butir – butir pengamatan yang mempunyai selisih ekstrim. Skor – skor yang mempunyai selisih ekstrim adalah pada butir pengamatan 1, 2, 5, 7, 8 9, 11, 12, dan 13 kemudian butir – butir tersebut didiskusikan bersama para rater untuk dicari penyebab dari perbedaan penskoran yang berbeda ekstrim. Setelah diketahui penyebab selisih ekstrim peneliti melakukan perbaikan terhadap instrumen dan menyepakati maksud yang diamati pada butir yang menimbulkan keraguan rater dalam memberi skor.

Hasil uji coba skala luas instrumen penilaian karakter tanggung jawab menunjukkan nilai koefisien korelasi antar kelas (*ICC*) tipe *consistent definition* dan *absolute agreement definition* meningkat untuk semua pasangan rater. Hasil uji reliabilitas dengan *ICC* dideskripsikan pada Tabel 2.

Tabel 2. Peningkatan Koefisien Reliabelitas Instrumen Pengamatan Keterampilan Memecahkan Masalah

Measures	<i>Interclass Correlation Coeffisien (ICC)</i>			
	Uji Coba Skala Kecil		Uji Coba Skala Luas	
	<i>Consistentensy</i>	<i>Absolute agreement</i>	<i>Consistentensy</i>	<i>Absolute agreement</i>
<i>Single Measure</i>	0,707	0,617	0,849	0,746
<i>Average Measures</i>	0.878	0,829	0,944	0,898

Keterangan (Utomo, 2012: 348):

1. Kriteria dengan tipe uji *consistency* minimal 0,70 ($\bar{r} > 0,70$)
2. Kriteria dengan tipe uji *Absolute agreement* minimal 0,60 ($\bar{r} > 0,60$)

Data tersebut menunjukkan bahwa pada saat uji skala kecil koefisien reliabelitasnya (*Single Measure*) instrumen penilaian keterampilan memecahkan masalah hanya sebesar 0,707 menggunakan analisis *ICC* tipe *Consistentensy definition*, sedangkan menggunakan analisis *ICC* tipe *Absolute agreement definition* sebesar 0,617. Namun demikian setelah dilakukan revisi dan kesepakatan yang dimaksud dalam pengamatan, koefisien meningkat menjadi 0,746 dalam tipe *Absolute agreement definition* dan 0,849 dalam tipe *Consistentensy definition*. Nilai koefisien reliabelitasnya tidak hanya meningkat namun juga melampaui dari batas yang ditetapkan (Utomo, 2012: 348), dengan demikian sifat reliabel pada instrumen penilain unjuk kerja pada keterampilan memecahkan masalah dipenuhi.

1.3. Instrumen Penilaian Unjuk Kerja Pada Kemampuan Memecahkan Masalah

Instrumen penilaian unjuk kerja pada kemampuan memecahkan masalah menggunakan bentuk tes uraian (Tim sergur, 2003: 15) dan Grounlund dalam Sukardi (2009: 94). Instrumen penilaian unjuk kerja pada kemampuan masalah dikembangkan menyesuaikan SK dan KD ada pada kurikulum KTSP. Instrumen yang baik (Azwar, 2011: 2) paling tidak harus memenuhi sifat Validitas dan sifat reliabelitas. Zulaiha (2007: 23) menambahkan soal bentuk uraian yang baik harus dianalisis secara kuantitatif meliputi daya pembeda dan tingkat kesukaran.

Kegiatan menganalisis data untuk menentukan nilai validitas item pada penelitian menggunakan uji korelasi *Pruduct Moment Pearson* (Alhusin, 2003: 336). Ternyata Butir soal yang mempunyai nilai koefisien korelasi di atas harga r tabel (0,279) dan merupakan butir soal yang mempunyai korelasi yang signifikan adalah 2, 4, 5, 6, 7 dan 8. Sedangkan untuk butir 1 dan 3 mempunyai nilai koefisien korelasi yang tidak signifikan. (Alhusin, 2003: 341)

Enam butir soal yang signifikan ini dianggap sebagai soal yang valid selebihnya dua butir soal dianggap tidak valid. Langkah selanjutnya, mengeluarkan dua butir soal yang tidak valid dari data semula dan kembali melakukan uji korelasi. Hasil uji korelasi nilai r untuk enam butir ternyata semuanya di atas harga r tabel 0,279 hal ini menunjukkan bahwa enam butir soal tersebut adalah benar valid dan dapat digunakan untuk mengukur unjuk kerja pada kemampuan memecahkan masalah.

Delapan butir soal kemampuan memecahkan masalah menggunakan analisis daya pembeda diperoleh nilai untuk butir 1, 2, 3, 4, 5, 6, 7, dan 8 berturut turut adalah 0.04, 0.27, 0.12, 0.62, 0.60, 0.39, 0.74, dan 0.45. Berdasarkan tabel kategori Mulyatiningsih (2013: 174) butir soal yang jelek dan harus dibuang adalah butir 1 dan butir 3.

Tingkat kesukaran untuk delapan butir soal kemampuan memecahkan masalah berturut – turut adalah 0.33, 0.86, 0.24, 0.69, 0.65, 0.57, 0.62, dan 0.48. Berdasarkan Mulyatiningsih (2013: 173) butir soal yang mempunyai tingkat kesukaran sulit adalah butir 1 dan butir 3.

Analisis kuantitatif terhadap butir soal uraian di atas memberi simpulan bahwa hanya ada enam butir soal yang mempunyai persyaratan instrumen penilaian yang baik menurut Azwar

(2011: 2) dan Zulaiha (2003: 23). Butir soal tersebut adalah butir soal nomor 2, 4, 5, 6, 7, dan 8. Dua butir soal lainnya gugur atau dibuang.

Pada uji coba skala diperluas pun enam butir soal kemampuan memecahkan masalah memenuhi persyaratan sebagai instrumen penilaian yang baik menurut Azwar (2011: 2) dan Zulaiha (2003: 23). Dengan demikian enam butir soal ini secara empirik membuktikan sebagai soal yang baik dan dapat mengukur kemampuan memecahkan masalah.

Simpulan

Pengembangan instrumen penilaian unjuk kerja dalam penelitian ini meliputi tiga ranah penilaian, yakni: (1) unjuk kerja pada ranah afektif mengukur karakter tanggung jawab, (2) unjuk kerja pada ranah psikomotor mengukur keterampilan memecahkan masalah, dan (3) unjuk kerja pada ranah kognitif mengukur kemampuan memecahkan masalah. Instrumen penilaian unjuk kerja ini diintegrasikan dengan proses pembelajaran matematika pada materi segitiga menggunakan model pembelajaran *problem solving* berbasis TIK.

Berdasarkan hasil – hasil penelitian yang disajikan dan dibahas, maka berikut ini adalah simpulan-simpulan pokok yang menjadi temuan dari penelitian ini.

1. Instrumen penilaian unjuk kerja pada pembelajaran matematika dengan model *problem solving* berbasis TIK terdiri atas sejumlah perangkat instrumen penilaian pada aspek afektif, kognitif dan psikomotorik. Perangkat instrumen telah terbukti memiliki tingkat validitas dan reliabilitas yang baik.
2. Penerapan instrumen penilaian unjuk kerja dalam pembelajaran matematika dengan model *problem solving* berbasis TIK memberikan informasi yang akurat tentang unjuk kerja siswa yang meliputi aspek sikap pada karakter tanggung jawab, aspek keterampilan memecahkan masalah, dan kemampuan memecahkan masalah.
3. Informasi hasil penilaian dari instrumen penilaian unjuk yang utamanya digunakan sebagai umpan balik kepada siswa dan refleksi bagi guru dalam usaha meningkatkan efektivitas proses pembelajaran.

Daftar Pustaka

- Alhusin, S. 2003. *Aplikasi Statistik Praktis dengan SPSS.10 for windows*. Yogyakarta: Graha Ilmu.
- Azwar, S. 2011. *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.
- Creswell, J.W. 2013. *Research Design: Pendekatan Kualitatif, Kuantitatif dan Mixed*. Terjemahan Fawaid, A. Yogyakarta: Pustaka Pelajar.
- Dwipurwani, O, dkk. 2009. “Penerapan Analisa Faktor dalam Membentuk Faktor Laten yang Mempengaruhi Prestasi Mahasiswa di Jurusan Matematika FMIPA Universitas Sriwijaya” *Jurnal Penelitian Sain*. No. 3A Vol. 12. Hal. 1 – 5
- Gareis, R.C. 2007. “Reclaiming an Important Teacher Competency: The Lost Art of Formative Assesment” *J Pers Eval Educ*. No. 20. Hal. 17–20
- Margaret, W. & Brown, T. 2012. “Affective Productions of Mathematical Experience” *Educ Stud Math* Vol. 80. Hal 185–199
- Mulyani, E. 2012. Pengembangan Model Penilaian Komprehensif Berbasis Proyek Pendidikan Kewirausahaan Terintegrasi di SMK. *Jurnal Penelitian dan Evaluasi UNY*. Tahun 16 No. 2. Hal 252 - 278.
- Mulyatiningsih, E. 2013. *Metode Penerapan Penelitian Terapan Bidang Pendidikan*. Bandung: Alfabeta
- Noviyanti, E. 2010. “Penilaian Kinerja (Performance Assessesment)”. <http://ekanoviyantifisika08unsri.blogspot.com/2010/10/penilaian-kinerja-performance.html>. (diunduh 5 Februari 2014)
- Praslova, L. 2010. “Adaptation of Kirkpatrick’s four level model of training kriteria to assessment of learning outcomes and program evaluation in Higher Education” *Educ Asse Eval Acc*. Vol. 22. Hal 215–225.
- Peter, E. et al, 2016, “‘Accepting Emotional Complexity’: A Socio-Constructivist Perspective On The Role Of Emotions In The Mathematics Classroom”, *Jurnal: Educational Studies in Mathematic*, Vol. 63 hal 193.
- Puskur. 2010. *Bahan Pelatihan Penguatan Metodologi Pembelajaran Berdasarkan Nilai – Nilai Budaya untuk Membentuk Daya Saing dan Karakter Bangsa*. Jakarta: Kementerian Pendidikan Nasional Badan Penelitian dan Pengembangan.
- Santoso, S., 2003. *Buku Latihan SPSS Statistik Multivariat*. Jakarta: PT Elex Media KomputindoKelompok Gramedia.
- Stanislaw et al, 2011, “Teaching Methods for Modelling Problems and Students’ Task-Specific Enjoyment, value, Interest and Self-Efficacy Expectations”, *Educ Stud Math* Vol. 79. Hal.215–237.
- Sudiyatno. 2010. “Pengembangan Model Penilaian Komprehensif Unjuk Kerja Peserta didik Pada Pembelajaran Berbasis Standar Kompetensi di SMK Teknologi Industri”. *Disertasi* . Yogyakarta: Program Pasca Sarjana Universitas Negeri Yogyakarta.
- Sugiyono. 2009. *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R & D*.

Bandung: Alfabeta

- Sukardi.2009. *Evaluasi Pendidikan Prinsip dan Operasionalnya*. Jakarta: Bumi Aksara
- Sukestiyarno. 2004. *Penerapan Strategi Berbasis Media dan Tehnologi dalam Mengajarkan Materi Matematika Perdana Sebagai Implementasi Kurikulum Berbasis Kompetensi*. Laporan Penelitian Due Like UNNES
- Tim Sergur Unnes, 2013a, “Elemen Perubahan Kurikulum 2013” dalam *Bahan Ajar Pendidikan dan Latihan Profesi Guru (PLPG) Sertifikasi Guru dalam Jabatan Bidang Studi Matematika SMP*, Semarang: Panitia Sertifikasi Guru Rayon 112 UNNES.
- Utomo, U. 2012, “Model Asesmen Kompetensi Guru Seni Musik dalam Perspektif Pelaksanaan Pembelajaran Berbasis *Action Learning*”. *Disertasi*. Yogyakarta: Program Pasca Sarjana Universitas Negeri Yogyakarta.
- Wardhani, S. 2004. *Penilaian Pembelajaran Matematika Berbasis Kompetensi*. Yogyakarta: Dirjen Pendidikan Dasar dan Menengah Pusat Pengembangan Penataran Guru (PPP) Matematika.
- Widhiarso, Wahyu. 2014. *Mengestimasi Reliabelitas*http://widhiarso.staff.ugm.ac.id/files/bab_2_estimasi_reliabilitas_via_spss.pdf (diunduh tanggal 23 Februari 2014)
- Widhiarso, Wahyu. 2014a. *Melibatkan Rater dalam Pengembangan Alat Ukur*.<http://widhiarso.staff.ugm.ac.id/files/Melibatkan%20Rater%20dalam%20Pengembangan%20Alat%20Ukur.pdf> (diunduh tanggal 23 Februari 2014)
- Widyaningsih, Y.S., Haryono dan Saputro, S. 2012. “Model Mfi Dan Pogil Ditinjau Dari Aktivitas Belajar Dan Kreativitas Peserta didik Terhadap Prestasi Belajar”.*JURNAL INKUIRI*. Vol 1. No 3. Hal 266-275.
- Zulaiha, R. 2007. *Analisis Butir Soal Secara Manual*.Jakarta: Pusat Penilaian Balitbang Depdiknas