

# **PENGEMBANGAN INSTRUMEN EVALUASI *TWO – TIER MULTIPLE CHOICE QUESTION* UNTUK MENGUKUR KETERAMPILAN BERPIKIR TINGKAT TINGGI**

**Mufida Nofiana<sup>1</sup>, Sajidan<sup>2</sup> dan Puguh Karyanto<sup>3</sup>**

**<sup>1</sup> Program Studi Pendidikan Sains, Program Pascasarjana  
Universitas Sebelas Maret Surakarta, 57126, Indonesia**

**<sup>2</sup> Program Studi Pendidikan Sains, Program Pascasarjana  
Universitas Sebelas Maret Surakarta, 57126, Indonesia**

**<sup>3</sup> Program Studi Pendidikan Sains, Program Pascasarjana  
Universitas Sebelas Maret Surakarta, 57126, Indonesia**

*Email: mufida.nofiana@yahoo.co.id*

Evaluasi merupakan alat yang digunakan untuk mengukur tujuan pembelajaran yang salah satunya adalah penguasaan keterampilan berpikir tingkat tinggi (*higher order thinking skills*). Penguasaan keterampilan berpikir tingkat tinggi membutuhkan kemampuan seperti menganalisis, mengevaluasi, dan menciptakan. Salah satu cara pengukuran keterampilan berpikir tingkat tinggi dapat dilakukan dengan instrumen evaluasi *two-tier multiple choice question*. Penelitian pengembangan yang dilakukan bertujuan untuk mengetahui (1) karakteristik instrumen evaluasi *two-tier multiple choice question* yang dikembangkan, (2) kelayakan instrumen evaluasi *two-tier multiple choice question* sebagai evaluasi formatif. Penelitian pengembangan menggunakan model *Research and Development* (R&D) mengacu pada Borg and Gall (1983) yang telah dimodifikasi. Sampel pengembangan meliputi 4 validator ahli, 20 siswa pada uji terbatas, dan 64 siswa pada uji lapangan. Analisis data dilakukan dengan rumus persentase. Hasil penelitian menunjukkan (1) karakteristik instrumen evaluasi *two-tier multiple choice question* yang dikembangkan antara lain mengacu pada indikator keterampilan berpikir tingkat tinggi Anderson dan Krathwohl (2001) meliputi menganalisis, mengevaluasi, dan menciptakan; memiliki validitas dengan interpretasi minimal “cukup”; dan reabilitas “tinggi” (2) kelayakan produk instrumen evaluasi dijamin melalui validitas isi dengan kategori “baik”; validitas konstruk dengan kategori “baik”; validitas butir soal dengan interpretasi minimal “cukup”, tingkat kesukaran soal dengan proporsi 15% mudah: 80% sedang: 5% sulit; daya pembeda soal dengan interpretasi minimal “cukup”, dan kepraktisan penggunaan dengan kategori “baik”.

**Kata kunci:** *two-tier multiple choice question*, keterampilan berpikir tingkat tinggi

## PENDAHULUAN

Keterampilan berpikir tingkat tinggi (*higher order thinking skills*) merupakan suatu keterampilan berpikir yang tidak hanya membutuhkan kemampuan mengingat, tetapi membutuhkan kemampuan lain yang lebih tinggi. Lewis dan Smith (1993) mendefinisikan keterampilan berpikir tingkat tinggi (*The Higher Order Thinking Skills*) sebagai keterampilan berpikir yang terjadi ketika seseorang mengambil informasi baru dan informasi yang sudah tersimpan dalam ingatannya, selanjutnya menghubungkan informasi tersebut dan menyampaikannya untuk mencapai tujuan atau jawaban yang dibutuhkan. King, *et al* (2010) mengatakan keterampilan berpikir tingkat tinggi pada siswa dapat diberdayakan dengan memberikan masalah yang tidak biasa dan tidak menentu seperti pertanyaan atau dilema, sehingga penerapan yang sukses dari kemampuan ini adalah ketika siswa berhasil menjelaskan, memutuskan, menunjukkan, dan menghasilkan penyelesaian masalah dalam konteks pengetahuan dan pengalaman.

Pentingnya penguasaan keterampilan berpikir tingkat tinggi terdapat dalam beberapa poin Standar Kompetensi Lulusan Sekolah Menengah. Poin yang diharapkan yaitu siswa dapat membangun dan menerapkan informasi atau pengetahuan secara logis, kritis, kreatif, dan inovatif; menunjukkan kemampuan berpikir logis, kritis, kreatif, dan inovatif dalam

pengambilan keputusan; serta menunjukkan kemampuan menganalisis dan memecahkan masalah kompleks (Permendiknas No 23 Tahun 2006). Pembelajaran Biologi merupakan pembelajaran sains yang memerlukan kegiatan penyelidikan atau eksperimen sebagai bagian dari kerja ilmiah. Kerja ilmiah menekankan peserta didik untuk berpikir kreatif, kritis, analitis, dan divergen (BSNP, 2006). Kemampuan peserta didik untuk berpikir kritis dan kreatif termasuk dalam bentuk keterampilan berpikir tingkat tinggi.

Konsep berpikir tingkat tinggi diturunkan dari Taksonomi Bloom. Sistem ini mengidentifikasi kemajuan yang hierarki dalam menggolongkan tingkatan proses berpikir menjadi tinggi dan rendah. Ada enam tingkatan taksonomi Bloom yakni: pengetahuan, pemahaman, aplikasi, sintesis, dan evaluasi. Tingkatan pertama dan kedua dari taksonomi Bloom dianggap sebagai kemampuan berpikir tingkat rendah, sedangkan empat tingkatan lainnya digolongkan sebagai keterampilan berpikir tingkat tinggi (Miller, 1990 dalam Ball dan Garton, 2005). Anderson dan Krathwohl (2001) telah merevisi penggunaan Taksonomi Bloom sebagai kerangka konseptual untuk penelitian keterampilan berpikir tingkat tinggi. Pohl (2000) mengungkapkan bahwa dalam Taksonomi Bloom revisi keterampilan yang melibatkan analisis, evaluasi, dan mencipta dianggap sebagai keterampilan berpikir tingkat tinggi. Anderson dan Krathwohl (2001) menyatakan bahwa indikator untuk mengukur keterampilan berpikir

tingkat tinggi meliputi menganalisis, mengevaluasi, dan menciptakan.

Instrumen evaluasi yang mengukur keterampilan berpikir tingkat tinggi dapat menggunakan berbagai tipe penilaian seperti *modified multiple choice*, konstruksi jawaban singkat, dan konstruksi jawaban panjang seperti yang telah dilakukan oleh Ramirez dan Ganaden (2008). Salah satu alternatif *Modified multiple choice* yang dapat digunakan untuk mengukur keterampilan berpikir tingkat tinggi adalah bentuk *two-tier multiple choice question* (pilihan ganda bertingkat). Bentuk soal *two-tier multiple choice question* dikembangkan oleh Treagust (2006). Treagust menggunakan bentuk soal *two-tier multiple choice question* untuk mendiagnosis kemampuan siswa memahami konsep IPA. Bentuk soal terdiri dari dua tingkatan soal, tingkatan pertama merupakan isi soal yang memiliki dua alternatif jawaban dan tingkatan kedua merupakan alasan jawaban yang dipilih atas dasar pilihan pertama. Pengembangan instrumen evaluasi *two-tier multiple choice question* dilakukan dengan mengaitkannya pada materi *kingdom plantae*.

Materi *Kingdom plantae* merupakan materi yang dekat dengan siswa. Contoh nyata dari materi *kingdom plantae* sering dijumpai di lingkungan sekitar, seharusnya siswa dapat menguasai materi tersebut dengan baik namun pada kenyataannya masih terdapat siswa yang tidak tuntas terutama pada soal-soal *kingdom plantae* yang menuntut keterampilan berpikir tingkat tinggi (Data UN

Tahun 2009 dan 2010). Penilaian keterampilan berpikir tingkat tinggi menggunakan *two-tier multiple choice question* pada materi *kingdom plantae* diharapkan mampu melatih siswa untuk memberdayakan keterampilan berpikir tingkat tinggi pada materi tersebut.

Halaydina dan Downing (1989) serta Treagust (2006) mengemukakan keunggulan bentuk soal *two-tier multiple choice question*, salah satunya digunakan untuk tujuan tes yang mengukur kemampuan kognitif siswa pada level yang lebih tinggi (*Higher Order Thinking*). Bentuk soal *two-tier multiple choice question* dapat digunakan untuk membantu menguji pemahaman siswa serta membantu mengidentifikasi miskonsepsi yang mungkin dimiliki oleh siswa. Cullinane (2011) mengemukakan penyertaan alasan pada tingkatan kedua dari bentuk soal *two-tier multiple choice question* dapat digunakan untuk meningkatkan keterampilan berpikir tingkat tinggi dan melihat kemampuan siswa dalam memberi alasan. Penyertaan alasan pada tingkatan kedua soal ini dapat digunakan untuk mengurangi terjadinya untung-untungan yang sering menjadi kelemahan dari bentuk soal pilihan ganda biasa. Penilaian soal yang objektif, mudah, dan cepat menjadi keunggulan *two-tier multiple choice question* dibandingkan dengan soal keterampilan berpikir tingkat tinggi yang lainnya contohnya soal essay. Kelemahan dari soal *two-tier multiple choice question* yaitu tidak mampu digunakan untuk mengukur

kemampuan verbal siswa seperti soal *essay*.

### Metode Penelitian

Penelitian dilaksanakan di SMA Negeri 3 Surakarta dan SMA Negeri 1 Gemolong. Sampel yang digunakan adalah siswa kelas X semester genap Tahun Pelajaran 2012/ 2013 berjumlah 149 siswa. Jenis penelitian yang digunakan adalah *Research and Development (R&D)*. Prosedur penelitian dimodifikasi dari model pengembangan Borg & Gall (1983) dan dilakukan hanya sampai pada tahap ketujuh.

Tahapan penelitian dan pengembangan meliputi 1) *research and information collecting*, yang dilakukan antara lain mengenali permasalahan yang ada di lapangan, analisis proses pembelajaran guru, analisis hasil UN, analisis kurikulum, analisis bank soal, dan studi pustaka; 2) *planning*, yang dilakukan antara lain menentukan Kompetensi Dasar materi yang akan dikembangkan, merumuskan tujuan pengembangan dan indikator keterampilan berpikir tingkat tinggi; 3) *develop preliminary from of product*, yang dilakukan antara lain menyiapkan materi pembelajaran, membuat kisi-kisi soal, mengembangkan produk awal instrumen evaluasi; 4) *preliminary field testing*, yang dilakukan antara lain validasi produk ke ahli dan guru senior, uji skala terbatas kepada 20 orang siswa yang terdiri dari 6 siswa untuk uji satu-satu dan 14 orang siswa untuk uji skala kecil; 5) *main product revision*, yang dilakukan antara lain

perbaikan sesuai dengan saran-saran dari hasil *preliminary field testing*; 6) *main field testing*, yang dilakukan antara lain menguji produk pengembangan dalam skala lebih luas pada 64 orang siswa di SMA Negeri 1 Gemolong untuk melihat validitas, reliabilitas, tingkat kesukaran dan daya beda soal; 7) *operational product revision*, yang dilakukan antara lain revisi produk berdasarkan saran-saran dari hasil *main field testing*. Hasil tahap *operational product revision* adalah produk final instrumen evaluasi *two-tier multiple choice question*.

Instrumen pengambilan data yang digunakan meliputi angket untuk analisis kebutuhan, lembar *check list* 8 SNP, lembar *check list* ketuntasan KD, dokumentasi silabus dan RPP guru, bank soal guru, lembar *check list* penilaian produk, dan lembar *check list* kepraktisan soal. Data analisis kebutuhan dianalisis dengan statistik deskriptif. Data penilaian ahli dan guru senior terhadap soal dianalisis dengan teknik deskriptif persentase (Purwanto, 2010). Analisis data dilakukan dengan cara menghitung skor yang dicapai dari seluruh aspek yang dinilai kemudian menghitungnya dengan rumus sebagai berikut:

$$N = \frac{k}{Nk} \times 100\%$$

Keterangan :

N : persentase kelayakan aspek

k : skor hasil pengumpulan data

Nk : skor maksimal (skor kriteria tertinggi x jumlah aspek x jumlah validator)

Skor yang diperoleh dibandingkan dengan Tabel 1 untuk mengetahui kriteria penilaian ahli.

Tabel 1. Kriteria Interpretasi Skor Validasi Ahli

Interval kriteria	Kriteria	Konversi
86 % N < 100%	Sangat baik	A
72 % N < 85%	Baik	B
58 % N < 71%	Cukup	C
44 % N < 57%	Kurang	D
N 44 %	Sangat kurang	E

Data penilaian siswa pada uji coba terbatas (uji coba satu-satu dan uji coba kelompok kecil) dianalisis dengan teknik deskriptif persentase (Purwanto, 2010). Perhitungan data yang dilakukan sama dengan perhitungan pada data validasi ahli dan guru senior. Pada uji coba lapangan terdapat dua jenis data, yakni data kualitatif dan data kuantitatif. Data kualitatif diperoleh dari data kepraktisan soal. Data kepraktisan soal dianalisis dengan teknik deskriptif persentase (Purwanto, 2010). Perhitungan data yang dilakukan sama dengan perhitungan pada data validasi ahli dan guru senior, sedangkan data kuantitatif diperoleh dari pengujian soal meliputi uji validitas, reliabilitas, daya beda, dan tingkat kesukaran soal.

Pengujian validitas dilakukan dengan menggunakan Microsoft excel 2007. Validitas instrumen tes tertulis dapat ditentukan dengan menggunakan rumus korelasi. Rumus korelasi yang

digunakan adalah rumus korelasional product moment dari Pearson. Pengujian reliabilitas dilakukan Microsoft Excel 2007. Rumus yang digunakan adalah rumus *Alpha-Cronbach*. Tingkat kesukaran soal dihitung melalui proporsi jawaban keseluruhan siswa yang menjawab benar pada soal tersebut. Daya pembeda dihitung melalui selisih jawaban antara proporsi kelompok tinggi yang menjawab benar dengan proporsi kelompok rendah yang menjawab benar.

Pengujian instrumen evaluasi pengembangan dilakukan pada siswa yang sebelumnya telah mendapat materi kingdom *plantae*. Hasil pengembangan instrumen diharapkan akan menghasilkan produk yang mampu memperbaiki kualitas soal pilihan ganda dan memperkaya khazanah soal-soal biologi di SMA khususnya pada materi kingdom *plantae*.

### Hasil Penelitian dan Pembahasan

Data yang diperoleh dalam penelitian pengembangan antara lain data analisis kebutuhan, data validasi ahli dan praktisi, data hasil uji coba terbatas dan data hasil uji coba lapangan. Data analisis kebutuhan meliputi tingkat pemenuhan standar nasional pendidikan (SNP) di SMA Negeri 1 Gemolong dan SMA Negeri 3 Surakarta, analisis bank soal biologi yang digunakan oleh guru di sekolah, dan wawancara.

Penerapan yang sukses dari kemampuan berpikir tingkat tinggi terjadi ketika siswa berhasil menjelaskan, memutuskan,

menunjukkan, dan menghasilkan penyelesaian masalah dalam konteks pengetahuan dan pengalaman (King, et.al, 2010). Pengembangan instrumen evaluasi untuk mengukur keterampilan berpikir tingkat tinggi belum banyak dilakukan oleh praktisi pendidikan. Penilaian formatif yang ada sekarang ini hanya sedikit memberikan kesempatan pada siswa untuk mengembangkan pengetahuan lebih mendalam (Cullinane, 2011). Instrumen evaluasi yang mampu mengukur keterampilan berpikir tingkat tinggi mempunyai beberapa indikator antara lain: cenderung kompleks, memiliki solusi yang mungkin lebih dari satu (*open-ended approach*), dan membutuhkan usaha untuk menemukan struktur dalam ketidakteraturan (Lewi, 2009).

Pengembangan instrumen evaluasi *Two-tier Multiple Choice Question (TT MCQ)* didasarkan pada teori perkembangan kognitif dari Piaget. Implikasi dari teori Piaget adalah instrumen yang dikembangkan disesuaikan dengan tingkat perkembangan kognitif siswa sehingga tidak terlalu sulit untuk dipahami. Bentuk instrumen yang dikembangkan sesuai dengan teori berpikir “John Dewey”. Implikasi teori Dewey dalam pengembangan instrumen evaluasi adalah soal yang diberikan berupa masalah yang bertujuan untuk merangsang siswa meningkatkan kemampuan berpikir yang tidak hanya sekedar menghafal. Indikator instrumen evaluasi yang dikembangkan sesuai dengan teori kognitif Bloom yang telah direvisi oleh Anderson dan Kratwohl (2001)

meliputi kemampuan menganalisis, mengevaluasi, dan menciptakan. Rangkuman hasil pengujian butir soal pengembangan disajikan pada Tabel 2.

Tabel 2 Rangkuman Hasil Pengujian Instrumen *TT-MCQ*

No soal	Keputusan	Interpretasi	Tingkat kesukaran	Daya pembeda
1	Valid setelah revisi	Cukup	Mudah	Cukup
2	Valid	Cukup	Sedang	Tinggi
3	Valid setelah revisi	Cukup	Sedang	Cukup
4	Valid	Cukup	Sedang	Tinggi
5	Valid	Cukup	Sulit	Tinggi
6	Valid setelah revisi	Cukup	Sedang	Tinggi
7	Valid setelah revisi	Cukup	Mudah	Tinggi
8	Valid	Tinggi	Sedang	Sangat tinggi
9	Valid	Cukup	Sedang	Tinggi
10	Valid setelah revisi	Cukup	Sedang	Cukup
11	Valid	Cukup	Sedang	Cukup
12	Valid	Cukup	Sedang	Tinggi
13	Valid	Cukup	Sedang	Tinggi
14	Valid	Tinggi	Sedang	Sangat tinggi
15	Valid	Cukup	Sedang	Tinggi
16	Valid	Cukup	Sedang	Tinggi
17	Valid	Cukup	Mudah	Cukup
18	Valid	Cukup	Sedang	Cukup
19	Valid	Tinggi	Sedang	Sangat tinggi
20	Valid	Cukup	Sedang	Tinggi

Mardapi (2008) mengemukakan waktu yang dibutuhkan untuk mengerjakan tes bentuk pilihan ganda adalah 2-3 menit untuk setiap butir tes. Sukardjo (2008) dalam Salirawati (2011) menyatakan ujian selama 90 menit jumlah butir tes pilihan ganda sekitar 20-30 soal, yang berarti setiap

butir soal dikerjakan selama 3-3,6 menit. Arikunto (2007) menyatakan bahwa alokasi waktu pengerjaan sebuah tes tergantung pada banyaknya butir tes dan bentuk soalnya. Berdasarkan kajian yang dilakukan maka waktu yang diberikan untuk mengerjakan soal *two-tier multiple choice question* adalah 60 menit untuk 20 soal yang diberikan, sehingga masing-masing butir soal dikerjakan selama 3 menit. Hasil penilaian kepraktisan instrumen evaluasi disajikan pada Tabel 3.

Tabel 3 Analisis Indikator Penilaian Kepraktisan Instrumen *TT-MCQ*

No	Indikator	Skor (%)	Kriteria
1	Biaya penyusunan tes terjangkau	62,50	Cukup
2	Waktu penyusunan tes tidak lebih dari 1 bulan	62,50	Cukup
3	Penyusunan tes dapat dilakukan guru biologi	87,50	Sangat baik
4	Penilaian tes mudah	87,50	Sangat baik
5	Mengolah hasil tes mudah	87,50	Baik
6	Pelaksanaan tes mudah	75,00	Baik
7	Waktu siswa untuk pelaksanaan tes di sekolah cukup	75,00	Baik
<b>Rata-rata</b>		<b>76,79</b>	<b>Baik</b>

Tabel 3 menunjukkan persentase rata-rata penilaian indikator kepraktisan instrumen evaluasi adalah 76,79% atau dinilai baik. Kepraktisan instrumen evaluasi adalah kemungkinan suatu instrumen evaluasi digunakan kembali oleh guru untuk

mengukur tujuan pembelajaran pada suatu saat nanti (Purwanto, 2010).

Temuan di lapangan menunjukkan bahwa soal hasil pengembangan dalam bentuk *two-tier multiple choice question* (pilihan ganda bertingkat) memiliki keunggulan dan kelemahan. Keunggulan soal *two-tier multiple choice question* antara lain jumlah materi yang dapat ditanyakan relatif lebih banyak dibandingkan dengan materi yang dicakup soal bentuk uraian; dapat mengukur jenjang kemampuan berpikir tingkat tinggi (analisis, evaluasi, mencipta) yang umumnya sulit dilakukan oleh soal pilihan ganda biasa; penskoran mudah, cepat, dan objektif; reliabilitas soal relatif lebih tinggi dibandingkan dengan soal uraian; dapat digunakan untuk mengukur kemampuan *problem solving*; dapat digunakan sebagai alat diagnosis pemahaman materi; dapat digunakan untuk mendeteksi miskonsepsi yang mungkin dimiliki siswa; dapat digunakan untuk mengetahui efektifitas pembelajaran yang dilakukan guru; peluang untuk menerka atau menembak jawaban lebih sedikit karena antara soal tingkat pertama dengan soal tingkat kedua saling berkait.

Kelemahan soal *two-tier multiple choice question* antara lain kurang dapat digunakan untuk mengukur kemampuan verbal; penyusunan soal yang baik memerlukan waktu yang relatif lama dibandingkan dengan bentuk soal yang lainnya; siswa belum terbiasa menggunakan soal dalam bentuk *two-tier multiple choice question*; dan guru belum pernah

menggunakan soal *two-tier multiple choice question*.

### **Kesimpulan dan Saran**

Kesimpulan dari penelitian pengembangan evaluasi antara lain:

1. Karakteristik instrumen evaluasi *two-tier multiple choice question* yang mengukur keterampilan berpikir tingkat tinggi antara lain dikembangkan berdasarkan indikator keterampilan berpikir tingkat tinggi dari Anderson dan Krathwohl (2001) meliputi keterampilan menganalisis, mengevaluasi, serta menciptakan, memiliki validitas dengan interpretasi minimal “cukup”, serta memiliki reabilitas yang tinggi.
2. Kelayakan produk instrumen evaluasi *two-tier multiple choice question* dijamin melalui validitas isi dengan kategori baik; validitas konstruk dengan kategori baik; validitas butir soal dengan interpretasi minimal cukup, tingkat kesukaran soal dengan proporsi 15% mudah: 80% sedang: 5% sulit, daya pembeda soal dengan interpretasi minimal “cukup”, serta tingkat kepraktisan soal dengan kategori baik.

### **Daftar Pustaka**

Afcariono, Muchamad. (2008). Penerapan Pembelajaran Berbasis Masalah untuk Meningkatkan Kemampuan Berpikir Siswa pada Mata Pelajaran Biologi. *Jurnal Pendidikan Inovatif*. 3(2): 65-68

- Anderson, L.W., & Krathwohl, D.R. (2001). *A Taxonomy for Learning, Teaching, and Assesing*. New York: Longman
- Arikunto, Suharsimi. (2011). *Dasardasar Evaluasi Pendidikan*. Jakarta : PT. Bumi Aksara
- Ball, Anna L., & Garton, Bryan L.. (2005). Modelling Higher Order Thinking: The Allignment Between Objectives, Classroom Discourse, and Assesment. *Journal of Agricultural Education*, Volume 46, Number 2, 2005.
- Borg, W.R., & Gall, M.D. (1983). *Educational Research An Introduction (4th Ed)*. White Plains: Logman Inc.
- Cullinane, Alison., & Liston, Maeve. (2011). *Two-tier Multiple Choice Question: An Alternative Method of Formatif Assessment for First Year Undergraduate Biology Students*. Limerick: National Center for Excellence In Mathematics and Education Science Teaching and Learning (NCE-MSTL).
- Halaydina, T.M., & Downing, S.M. (1989). *A Taxonomy of Multiple Choice Item Writing Rules*. *Applied Measurements In Education*, 2(1), 37-50.
- King, JF; Goodson, Ludwika., & Rohani, Faranak. (2010). *Higher Order Thinking Skills, Definition, Teaching Strategis, Assesment*. A Publication of The Educational Services Program. Tersedia di [www.Cala.fsu.edu](http://www.Cala.fsu.edu)
- Lewis, A & Smith, D. (1993). *Defining Higher Order Thinking*. *Theory Into Practice*, 32(3), 131-137

- Mardapi, Djemari. (2008). *Teknik Penyusunan Instrumen Tes dan Nontes*. Yogyakarta: Mitra Cendekia Yogyakarta Press.
- Permendiknas No 20. (2007). *Standar Penilaian Pendidikan*. Jakarta: Depdiknas.
- Permendiknas No 23. (2006). *Standar Kompetensi Lulusan untuk Satuan Pendidikan Dasar dan Menengah*. Jakarta: Depdiknas.
- Pohl. (2002). *Learning Thinking to learn*. tersedia di [www.purdue.edu/geri](http://www.purdue.edu/geri)
- Purwanto, Ngalim. (2010). *Prinsip-prinsip dan Teknik Evaluasi Pengajaran*. Bandung: PT Remaja Rosdakarya.
- Ramirez, Rachel Patricia B., & Ganaden, Mildred S.. (2006). Creative Activities and Students' Higher Order Thinking Skills. *Journal of Education Quarterly*, December 2008, vol 66 (1), 22-23.
- Richmond, Jonathan E.D. (2007). Bringing Critical Thinking to The Education of Developing Country Professionals. *International Education Journal*, 2007, 8(1), 1-29.
- Sagala, Syaiful. (2009). *Konsep dan Makna Pembelajaran: untuk Membantu Memecahkan Masalah Belajar dan Mengajar*. Bandung: Alfabeta.
- Sajidan. (2012). Penerapan Model Pengembangan Mutu Pendidikan dalam Rangka Peningkatan Kompetensi Guru SMA Melalui Pengembangan *Subject Specific Paedagogy (SSP)*. *Draft Artikel Penelitian*. Surakarta: Universitas Sebelas Maret Surakarta.
- Salirawati, Das. (2011). Pengembangan Instrumen Pendeteksi Miskonsepsi Kesetimbangan Kimia pada Peserta Didik SMA. *Jurnal Pendidikan dan Evaluasi Pendidikan* Tahun 15, Nomor 2, 2011.
- Suryabrata, Sumadi. (2005). *Psikologi pendidikan*. Bandung: Rajawali Pers.
- Treagust, David F. (2006). Diagnostic Assesment In Science as A Means to Improving Teaching, Learning, and Retention. *UniServe Science Assesment Symposium Proceedings*. The University of Sydney, 28 September 2006.
- Weiss, Renée E. (2003). Designing Problems to Promote Higher Order Thinking. *New Direction for Teaching and Learning*, No 95, Fall 2003.