1) (4.3) $v_\pi(s) = \mathbb{E}_\pi [R_{t+1} + \gamma v_\pi(S_{t+1}) \mid S_t = s]$
$q_\pi(s, a) = \mathbb{E}_\pi [R_{t+1} + \gamma G_{t+1} \mid S_t = s, A_t = a]$

(4.4) $v_\pi(s) = \sum_a \pi(a|s) \sum_{s', r} p(s', r|s,a) [r + \gamma v_\pi(s')]$
$q_\pi(s, a) = \sum_{s', r} p(s', r|s, a) \sum_{a'} \pi(a'|s') [r + \gamma q_\pi(s', a')]$

(4.5) $v_{k+1}(s) = \sum_a \pi(a|s) \sum_{s', r} p(s', r|s,a) [r + \gamma v_k(s')]$
$q_{k+1}(s, a) = \sum_{s', r} p(s', r|s, a) \sum_{a'} \pi(a'|s') [r + \gamma q_k(s', a')]$

**Initialization:**
$Q(s, a) \in \mathbb{R}$ and $\pi(s) \in A(s)$ arbitrarily for all $s \in S$

**Policy Evaluation:**
Repeat
    $\Delta \leftarrow 0$
    For each $s \in S$:
        $q \leftarrow Q(s, \pi(s))$
        $Q(s, \pi(s)) \leftarrow \sum_{s', r} p(s', r|s, \pi(s)) \sum_{\pi(s)'} \pi(\pi(s)'|s') [r + \gamma q_\pi(s', \pi(s)')]$
        $\Delta \leftarrow \max(\Delta, |q - Q(s, \pi(s)))$
until $\Delta < \theta$ (a small positive number)

**Policy Improvement:**
policy-stable $\leftarrow$ true
For each $s \in S$:
    old-action $\leftarrow \pi(s)$
    $\pi(s) \leftarrow \text{argmax}_{a'} \sum_{s', r} p(s', r|s, a) \sum_{a'} \pi(a'|s') [r + \gamma q_\pi(s', a')]$
    If old-action $\neq \pi(s)$, then policy-stable $\leftarrow$ false
If policy-stable, then stop and return $Q \approx q_*$ and $\pi \approx \pi_*$; else go to **Policy Evaluation**